Spatio-temporal methods for EEG analysis in cognitive neuroscience

Poulsen, Andreas Trier

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Spatio-temporal methods for EEG analysis in cognitive neuroscience

Andreas Trier Poulsen

Kongens Lyngby 2018
PHD-2018-473

DTU Compute
Department of Applied Mathematics and Computer Science

Spatio-temporal methods for EEG analysis in cognitive neuroscience
PHD-2018-473

Andreas Trier Poulsen
atpo@dtu.dk

Technical University of Denmark
Department of Applied Mathematics and Computer Science

DTU Compute
Richard Petersens Plads
Building 324
2800 Kgs. Lyngby
DENMARK

Tel. +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk
PHD: ISSN 0909-3192

Ask not what you can do for machine learning.
Ask what machine learning can do for you.

Summary

Electroencephalography (EEG) records electrical activity from the brain by measuring the resulting potential differences across the scalp. It has a long tradition in both a clinical and neuroscientific setting, and recently it has also started being used for consumer-oriented applications.

While EEG can be a useful tool, it can be difficult to decipher information from its raw signals. In this thesis I will present three projects with the common goal of analysing EEG in ways that both extract meaningful information and visualise it in intuitive ways.

The first project describes how we took neuroscience out of the laboratory and into the classroom. We reproduced an attention-tracking paradigm in a classroom and simultaneously recorded the neural activity of up to nine people. We had a focus on using equipment that was wireless and portable as well being relatively low-cost and computational methods in a setup that is feasible to extend into everyday scenarios.

The second project revolved around creating a toolbox for the research field of microstate analysis, with a focus on open access and transparency of the applied methods. The toolbox is followed by a methodological guide that reviews the most commonly applied algorithms in microstate analysis.

In the final project I investigated the feasibility of using the complexity of EEG as a neural marker of conscious processing. This project spans two studies investigating the capability of EEG complexity in two different scenarios; while people are sleeping, and while navigating a helicopter simulator.

Resumé

Elektroencefalografi (EEG) optager elektrisk aktivitet fra hjernen ved at måle forskelle elektrisk potentiale på tværs af hovedbunden. EEG har en lang tradition både i klinisk og neurovidenskabelig omgivelser, og for nylig er den også begyndt at blive brugt til forbrugerorienterede applikationer.

Selvom EEG kan være et nyttigt værktøj, kan det være svært at afkode oplysninger fra dets rå signaler. I denne afhandling vil jeg præsentere tre projekter med det fælles mål at analysere EEG på måder, der både udtrækker meningsfuld information og visualiserer det på intuitive måder.

Det første projekt beskriver, hvordan vi tog neurovidenskab ud af laboratoriet og ind i klasseværelset. Vi reproducerede et paradigme måling af i et klasseværelse, og optog simultant den neurale aktivitet af op til ni personer. Vi havde fokus på at bruge udstyr, der var trådløst og bærbart, samt havde relativt lave omkostninger. Vi havde også fokus på at bruge metoder der ikke kræver beregningsmæssig ressourcer, med henblik på muligheder for at udvide det til hverdagens situationer.

Det andet projekt drejede sig om at skabe en toolbox til forskningsområdet for *microstate analysis* med fokus på åben adgang og gennemsigtighed i de anvendte metoder. Vores toolbox akkompagneres af en metodologisk vejledning, der gennemgår de mest anvendte algoritmer i *microstate analysis*.

I det sidste projekt undersøgte jeg muligheden for at bruge kompleksiteten af EEG som en neural markør for bevidst processering. Dette projekt spænder over to forskellige forsøg, der undersøger brugsmulighederne af EEG-kompleksitet i to forskellige scenarier; mens folk sover, og mens man navigerer en helikopter i en simulator.

Preface

This thesis was prepared at DTU Compute in the section of Cognitive Systems in fulfilment of the requirements for acquiring a PhD degree.

The thesis includes of a summary report of the experiments, the underlying theory, key findings and perspectives on these. The thesis, furthermore, contains published articles and draft manuscripts produced during the PhD. The thesis work was performed in the period December 15, 2015, to April 10, 2018.

Kongens Lyngby, April 10th, 2018.
Andreas Trier Poulsen

Acknowledgements

This PhD has been a journey and an adventure in both a figurative and literal sense, that has brought me into contact with many interesting people from many backgrounds and places. Their acquaintances has shaped both me and my work, and I owe thanks to them all.

Thank you to all of the people at Cognitive Systems, from my fellow students to the administrative staff to the professors. Thank you for being my work family. To the people at École Normale Supérieure, LSCP, for welcoming me warmly every time I popped through the doors at Châtelet and enthusiastically showing your newest progress.

For many interesting neuroscientific discussions over coffee or beer, I owe thanks to Ivana Konvalinka, Tobias Andersen, Sirin W. Gangstad, Frans Zdyb, Alma Lindborg, Martin Axelsen, Nikolaj Bak, Melissa Larsen, Sofie Therese Hansen, Laura Frølich, Albert Vilamala, Konrad Stanek, Søren Føns Vind Nielsen, Søren Hauberg, Per Bækgaard, Alya Vlassova, Hao Zhang, Quentin Fresnel, Oscar Lozano, Nicolas Barascud, Benjamin Rebouillat, Matthieu Koroma, Leonardo Barbosa, Jean-maurice Leonetti, and Sofie Gelskov.

To my collaborators; Simon Kamron, Thomas Andrillon, Jacek Dmochowski, Lucas C. Parra, Nicolas Langer and Andreas Pedroni, thank you for your patience.

A special thank you to my supervisors Lars Kai Hansen and Sid Kouider. I have been blessed with the guidance of not only one but two brilliant and accomplished researchers, who are experts in distinct fields that overlap in the same area as the topics presented in this thesis. Describing all the ways they have influenced and helped me become the scientist I am today, would require an entire chapter by itself. Lars and Sid, thank you for all of your time, and for continuing to teach and challenge me. Hopefully, we will have more journeys together in the future.

Friends and family in general also deserve a significant amount of gratitude for enduring the verbosity of my answers when asked a simple; *how's work?*

I am thankful for the financial support that I received for my external stays in Paris from the foundations of Stibo, Oticon, Idas & Berg-Nielsen, and Marie & M.B. Richter. I additionally wish to thank the Otto Mønsted Foundation for supporting me in conference participation.

Finally, I dedicate this work to the two women in my life. To Sofie, for being my companion through life. For keeping me grounded when everything is going too fast and for picking me up when I fall. To Elin, for showing me joys in life that I didn't know existed.

Contributions

Articles in peer-reviewed journals

- Andrillon, T., Poulsen, A. T., Hansen, L. K., Léger, D., Kouider, S. (2016). *Neural Markers of Responsiveness to the Environment in Human Sleep.* The Journal of Neuroscience, 36(24), 6583–6596.
- Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., Hansen, L. K. (2017). *EEG in the classroom: Synchronised neural recordings during video presentation.* Scientific Reports, 7, 1–9.

Articles for open access (not peer-reviewed)

- Poulsen, A. T., Pedroni, A., Langer, N., Hansen, L. K. (2018). *Microstate EEGlab toolbox: An introductory guide.* bioRxiv, 1–30.

Manuscripts in progress

- Poulsen, A. T., Leonetti, J., Hansen, L. K., Kouider, S. (In progress). *EEG complexity tracks user performance and difficulty in helicopter simulator.*
- Poulsen, A. T., Leonetti, J., Hansen, L. K., Kouider, S. (Submitted). *Tracking difficulty in a helicopter simulator: EEG complexity as a marker for mental workload.* Conference abstract.

Software

- Microstate analysis Matlab toolbox: A plugin for EEGlab. https://archive.compute.dtu.dk/files/public/users/atpo/Microstate

Contents

CHAPTER 1
Introduction

This thesis covers projects in the intersection of two disciplines; machine learning and cognitive neuroscience. The digital technological advances of the last couple of decades have made it possible for many sub-fields of research to emerge in this intersection. In general the machine learning/cognitive neuroscience intersection can be divided into three sub-fields:

1. Creating theoretical models for how the brain functions based on mathematical and statistical principles.
2. Using neuroscience to inspire the formulation of new machine learning methods.
3. Applying existing machine learning methods to neuroscientific data.

Though there is an overlap between these definitions, the projects presented here mainly resides in the latter field.

1.1 EEG as a tool in cognitive neuroscience

A common tool in cognitive neuroscience is the electroencephalography (EEG), which records electrical activity from the brain by measuring the resulting potential differences across the scalp.

Clinically, EEG can be used to help diagnose afflictions such as epilepsy, sleep-disorders, or brain damage. The clinical use of EEG is often preceded by research into how EEG can be interpreted. In cognitive neuroscience, EEG can help investigate how our brain works such as defining what consciousness and unconsciousness is. For example, by looking at the strength of potentials at specific locations and specific times in the EEG, it is possible to examine how the brain can analyse incoming visual or audio information, in both a conscious or unconscious manner and how these differ (Dehaene and Changeux, 2011). This information can then be extended to investigate when the same processes are present in non-verbal infants and thereby infer presence of consciousness in them (Kouider et al., 2013).

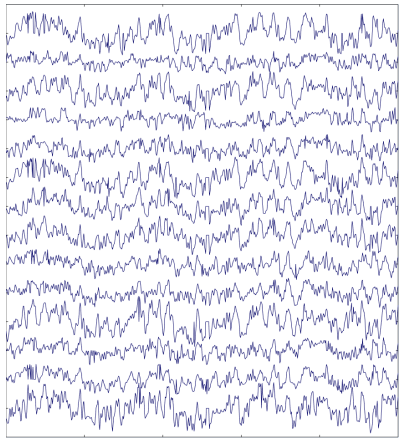


Figure 1.1: Five seconds of EEG recorded from 14 locations on the scalp.

While EEG is a useful tool, it can be difficult to decipher information from the raw signals. Figure 1.1 shows an example of raw EEG signals that, to the untrained eye, might seem like noise. For trained physicians it takes several years to learn to decode EEG, and even then there are limits to the amount of information they can read from it.

Like every other tool, EEG has its strengths and weaknesses. EEG has a high temporal resolution with sampling frequencies that can lie above 1,000 Hz and signals of interest usually lying in the 0.5 - 100 Hz area (Lopes da Silva, 2013). Compared to other neuroimaging techniques, EEG is relatively inexpensive, unobtrusive and portable. Finally, the biophysical foundation of the recorded signals is well understood (Buzsáki et al., 2012). These attributes makes EEG a good tool for a variety of clinical and experimental applications and a good candidate for bringing neuroscientific measurements out of the laboratory and into our everyday life.

Regarding the weaknesses of EEG, especially two aspects make it hard to decipher visually. EEG usually has a low signal-to-noise ratio since the signals are attenuated and distorted by the different layers between the electrical currents in the brain and the surface-electrodes on the scalp. This attenuation means that the recorded neural signals are so weak that they are easily contaminated by noisy signals from our surroundings. A second issue is that it is difficult to visualise how spatial patterns of neural activity across the scalp evolves in time.

Since the 1960s and 1970s, the digital age has helped alleviate some of these challenges of EEG analysis. By designing experiments to be time-locked to a short stimulus and repeating it multiple times, it became possible to average EEG to attenuate noisy signals unrelated to the experiment, and thereby enhance the event-related potentials (Hari and Puce, 2017). The advent of computers also made it possible to visualise the spatial distribution of potentials over the scalp for a given time in the EEG.

These breakthroughs helped improve the usefulness of EEG, but also restricted much

of the experiments to be short and repeatable. Additionally, analysis of temporal waveforms was often restricted to one EEG channel at a time, followed by spatial topographies at time points of interest. These constraints were not because there was no more information to be found by analysing the EEG altogether, but because there is a limit to how complex data can be, if it is a human who should find a pattern in it.

In the current phase of our digital revolution it is now possible to do high amounts of computation in a short time, which makes it possible to implement machine learning models to capture patterns in complex neuroscientific data, such as EEG. Patterns that would be hard for humans to detect. In other words, by using machine learning we can improve the efficacy of EEG as a tool in cognitive neuroscience.

1.2 Enhancing neuroscientific analysis with machine learning

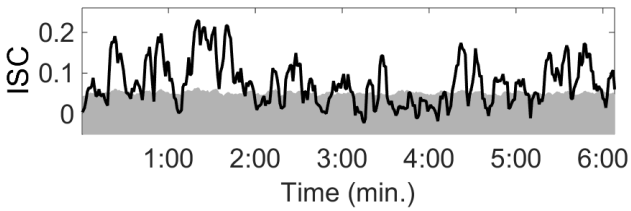
Humans excel in finding patterns in 2D images, but struggle when the dimensionality increases. Machine learning, when wielded correctly, is a powerful tool in finding patterns in data of both high and low dimensionality.

In these modern times we experience machine learning in many ways and many times during our day. How we use our smartphones is a good example of how machine learning finds patterns in information in order to help us. Our phones use machine learning when they find patterns in the sounds we make during our speech to understand what we are saying; when we ask our phones to find the fastest way to the cinema; when social media estimates which story to put at the top of our news feed; or when the patterns in our finger prints or facial structure are used to identify us.

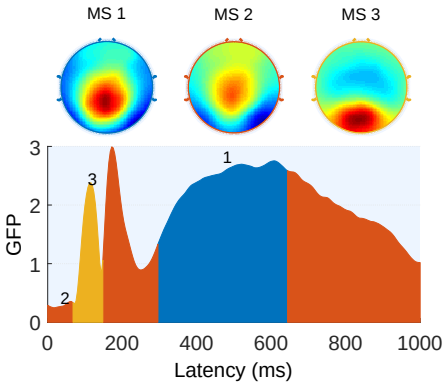
In my PhD I have sought to use methods that are both able to increase the amount of the information gained from EEG recordings, and able to visualise this information in ways that are more easy to understand intuitively. Figure 1.2 shows three examples of visualisations of EEG, from methods used in this thesis.

Some of the research results presented in this thesis might appear impractical to apply in real life scenarios. However, apart from the fact that most research needs technological refinement to bridge the gap into the hands of consumers, the results presented here must also be seen in the light of a recent revolution in the size and portability of EEG recording devices. It seems unrealistic that students would don large caps filled with gel before a lecture, but envision a future where EEG can be recorded from small devices that fit in the ear (Kidmose et al., 2013) or via thin electrode-patches applied to the skin (Norton et al., 2015).

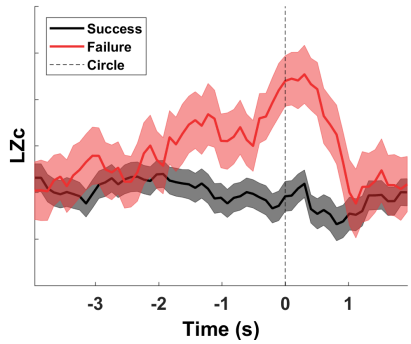
To summarise; the general goal of my PhD is to apply machine learning to neuroscientific data in order to obtain information that is salient in describing its origin, but at the same time is still easy to interpret.



(a) Inter-subject correlation.



(b) Microstate analysis.



(c) EEG complexity.

Figure 1.2: Examples of visualisations of EEG analysis. (a): The inter-subject correlation between subjects viewing the same film, with peaks coinciding with engaging scenes. The grey area indicate chance levels for ISC ($p > 0.01$, uncorrected). This is covered in chapter 2. (b): Microstate analysis of an event-related potential stemming from subjects viewing images with faces on them. The global field power (GFP) is calculated across all electrodes.The colour denotes which microstate the EEG samples has been assigned to with their corresponding spatial map shown above. This is covered in chapter 3. (c): EEG complexity for a subject flying a helicopter simulator. The complexity of each trial has been time-locked to an event and averaged in two groups based on whether the pilot succeeded or failed the trial. This is covered in chapter 4.

1.3 Spatio-temporal models

Sometimes terminology can vary between different research fields, and the same word can give different associations depending on which field a researcher is affiliated with. For this reason I will briefly clarify in what sense the term *spatio-temporal models* is used in this thesis.

When I define the different analysis methods presented throughout this thesis as spatio-temporal models, I do so in the literal interpretation of the words. They are models that in different ways incorporate information of both time and space. For EEG, the spatial component of the information stems from the spatial distribution of the electrodes on the two-dimensional surface of the scalp, and the underlying three-dimensional activity in the cortex it represents.

In the next section, I will give a short overview of the topic of each project, and how the employed methods can be considered to be spatio-temporal.

1.4 Motivation

The overarching goal of this thesis is to use machine learning methods to increase the information gained from recorded EEG. I have two motivations for pursuing this increase of extracted information.

One motivation is to apply the methods to the emerging field of wearable EEG. Neurotechnology develops fast and holds great societal promises in personalised medicine and patient monitoring as well as for neuroprosthetics such as hearing aids or brain-computer-interfaces. There is also a range of possible applications in more consumer-related areas, e.g. neuromarketing or future brain-based augmented reality interfaces.

These applications could be based on permanent or long-term mental state decoding, i.e. estimating the perceptual or cognitive state of the human brain, whether it is reflecting endogenous or external processes. With future EEG devices allowing prolonged recordings in a natural everyday setting, a deeper understanding might be obtained regarding the causal relationship between the frequency and times of occurrence of different brain states and many important aspects of life such as mental fitness (e.g. feelings of being energetic, stressed, unfocused) or mental focus (e.g. lectures, mental tasks, emotional attention).

A second motivation is to improve the insights gained from experiments in the field of cognitive neuroscience. Machine learning can either boost the results of classic experiments or be taken into account from the start to create novel experimental paradigms. For example, machine learning has been used in the study of variability in audiovisual integration (Keil et al., 2012), track representational similarity in object recognition (Cichy et al., 2014) and relate it to semantic processing using deep learning models (Khaligh-Razavi and Kriegeskorte, 2014). Machine learning has also been used for online decoded neurofeedback (Cortese et al., 2016), studying the neural coupling in interactions between infants and adults (Leong et al., 2017), and the unconscious maintenance of subliminal stimuli (King et al., 2016).

Most of these examples show an applied aspect of machine learning, where machine learning is used as a tool to solve specific challenges, rather than studying the performance of different models and comparing them to other models. It is examples like these that have influenced me to gravitate from asking *"what I can do for the field of machine learning?"* to *"what can machine learning do for me and what I want to achieve?"*

1.5 Contributions

In my PhD project, I have been blessed by not only having scientific freedom through a broad topic description, but also by having two inspiring supervisors. With a

common interest in applying machine learning to neuroscientific data, they approach the inter-disciplinary field with different backgrounds, and have challenged me to improve from both a machine learning and a cognitive aspect. This has altogether meant that I have had the opportunity to work on many different projects instead of having a single goal for the entire PhD project.

My three years as a PhD-student have therefore been less like an essay over a topic and more like a collection of short stories over a theme. This is also reflected in the structure of the resulting thesis, that you are about to read.

I have taken a pragmatic, and somewhat unorthodox, approach to the structure of this thesis, with the aim of keeping verbosity at a minimum. Each chapter will serve as a short walk-through of the concept and motivation of a project, as well as present selected relevant results. I will therefore refer to their associated articles for the full scientific background of the results and a more formal presentation of them. Since each chapter presents and summarises the results independently of each other, there is no extra chapter dedicated to the re-summarising the results. Also, this thesis will not contain a large theoretical chapter on the physics and physiology behind EEG. This information has already been written many times before and I instead refer to people who have explained it much better than I would be able to, such as Hari and Puce (2017).

Though each project contains spatio-temporal models and cognitive neuroscience, they are distinct both in the methods employed and how involved I have been in conducting the experiments:

Chapter 2 - *Classroom EEG*
This chapter describes how we took neuroscience out of the laboratory and into the classroom. We reproduced a previously published paradigm tracking attention with inter-subject correlation (ISC, Dmochowski et al., 2012), using low-cost portable EEG recording devices. We both reproduced the original offline-synchrony paradigm and extended it by recording EEG from nine subjects simultaneously. Furthermore, we investigated the neural origin of ISC.
In order to calculate ISC we used correlated component analysis. This method can be considered spatio-temporal in the sense that it estimates spatial components so they maximise the temporal correlation between components from different subjects. Additionally, in our study we calculated the ISC in a moving window of 5 seconds with a 80% overlap to obtain a one-second temporal resolution of the synchronised spatial activity.
The presented work is covered in the article available in appendix A.

Chapter 3 - *Toolbox for microstate analysis*
This chapter describes how we developed a toolbox for microstate analysis and a review of the methodology employed in the field. Since we focused only on the methodology, this chapter does not present new data nor go into the possible cognitive interpretations from microstate analysis.
The concept of microstate analysis is to cluster EEG into microstate clusters depending

on the similarity of their spatial topography. This way the EEG can go from being a multi-dimensional time series to be represented by a single-dimensional time series of class labels. Once each EEG time sample has been assigned to a (spatial) microstate class, the temporal dynamics of the changes in states can be analysed such that both the temporal and spatial aspect of EEG is taken into account.
The presented work is covered in the article available in appendix B.

Chapter 4 - *EEG complexity as a neural marker*
This chapter contains two projects investigating the properties of using EEG complexity as a neural marker. The complexity of the EEG is based on an algorithm for evaluating the complexity or randomness of finite sequences (Lempel and Ziv, 1976). The first project covers its usefulness for capturing characteristics of EEG recorded during sleep. The second project investigates if complexity can be used in a learning paradigm for subjects using a helicopter simulator.
In order to calculate complexity, the EEG first needs to be converted to a binary string. It is this conversion that can include spatial information to the method, and there are several approaches to do this. I use short windows (e.g. 1,500 ms), which is converted to a string by concatenating the binarised EEG observation-by-observation. It is the concatenation of the multi-dimensional channel activity of each observations that, in principle, retains the spatial information.
The presented work is covered in the article and manuscript drafts available in appendices C, D and E.

CHAPTER 2

Inter-subject correlation in a classroom

This chapter covers the concept of using inter-subject correlation (ISC) from EEG as a marker of attention. The chapter centres on how we showed that it is feasible to take the method out of the laboratory and perform it in a classroom using consumer-level devices to wirelessly record EEG on tablets (Poulsen et al., 2017). The article is available to the reader in appendix A.

Parts of the work presented in this article was performed prior to the start of this PhD project. The experiments were conducted in collaboration with Simon Kamronn for our master thesis (Kamronn and Poulsen, 2013). However, due to time constraints we were unable to spend enough time on the preprocessing and analysis of the recorded EEG, resulting in mixed results. Therefore, the preprocessing and analysis was redone from scratch for our article, and we additionally added new analyses as well as started a collaboration with Lucas Parra and Jacek Dmochowski.

2.1 Motivation

What if we could measure attention? Imagine if during a lecture a professor could get feedback, when the neural activity of her students became less synchronised, so she knew when to make a joke or have a pop-quiz to catch their attention. What if the students could get feedback, such that when their synchronisation with their peers had a relative drop, their phone vibrated? Either when participating in an auditorium or in an online course.

We are not there yet, but the progress in portable EEG devices and other bio-wearables, means that we are close to making it technologically feasible to record the data needed in real-time. The question is, are we able extract measures of attention from the data?

There are currently several approaches to solving the problem of tracking user attention through their neural responses. Here I will focus on using ISC to estimate

synchronisation between multiple users experiencing the same stimuli to track their engagement, as well as describe how we sought to bring the science behind the method out of the laboratory and into a real world scenario.

In this chapter I will start by outlining some background on the literature of measuring ISC during film viewing and correlated component analysis (CorrCA), which is used to calculate ISC for EEG. Furthermore, I will give an overview and a discussion of the results of Poulsen et al. (2017), concluding with a brief review of current work on the relationship between ISC and attention.

I will attempt to give the reader an overview of the field, as well as our contributions to it, and refer to our article and the cited references for a more in-depth review of the research.

2.2 Background

2.2.1 Neurocinematics: Watching films in neuroscience

Most neuroscientific experiments are constructed in a discrete manner, with short stimuli that can be repeated many times. By time-locking the stimuli and averaging over trials to obtain event-related potentials (ERPs), the signal-to-noise ratio can be improved and the neural responses of interest can be studied better. This is a very effective method to deal with the high levels of noise present in most brain imaging modalities, but the format also sets some constraints on the experimental paradigms.

In Hasson et al. (2004), the authors wanted to investigate neural responses under more "natural" conditions, namely while subjects viewed films, using functional magnetic resonance imaging (fMRI). Since films are continuous, it wasn't possible to average over multiple trials in the manner of ERP analysis and they instead opted for using ISC. By correlating the continuous responses of multiple subjects experiencing the same film it was possible to find times of synchronisation that "consisted of a widespread cortical activation pattern correlated with emotionally arousing scenes and regionally selective components".

In Hasson et al. (2008), they sought to expand on these results and investigate how ISC related to subject attention. They did this by introducing scrambled versions of the shown films, such that the subjects were still exposed to the same audio-visual stimulus, as subjects viewing the original films. The reasoning behind this is that by changing the order of the scenes the narrative will be disrupted, become less interesting, and thereby elicit a lower attentional effect in the viewers.

The authors found that a short by Alfred Hitchcock (*Bang! You're dead*, 1961) elicited the strongest and most spatially widespread synchronisation. This they speculated could be "neuroscientific evidence for his notoriously famous ability to master and manipulate viewers' minds". This lead Naci et al. (2015) to use ISC and Hitchcock to differentiate between degrees of awareness in comatose patients, by investigating their conscious experiences of the film.

These studies were done using fMRI, which may have a high spatial resolution (at the cost of a low temporal sampling frequency), but have considerable practical

constraints. fMRI is expensive to record and require the user to lie in a tube surrounded by super-cooled helium in a shielded room. This creates limitations with respect to giving viewers a cinema-like experience in a natural environment (at least with the present technology). Though the current laboratory-grade equipment used for EEG experiments also has some practical limitations to recording under natural and unobtrusive conditions, the advent of portable EEG devices and new types of electrodes has even greater promise for bringing this research out of the laboratory and into the real world.

In Dmochowski et al. (2012), Lucas Parra's group investigated if the link between ISC and attention (or user engagement) could be transferred to EEG analysis. To this end, the authors introduced a new analysis method based on canonical correlation analysis (see section 2.2.2). The authors used two of the same film clips as in the studies by Uri Hasson, including the Hitchcock short. Focusing on the three components with the highest ISC, they found that peaks of ISC coincided with scenes with high levels of suspense, tension, or surprise. They furthermore found that the ISC was lower when they sought to reduce engagement with the stimuli, either by scrambling the scenes, showing the film a second time, or showing an uneventful control film (natural outdoor scene of a college campus). Again the Hitchcock short proved to elicit the strongest ISC.

In Dmochowski et al. (2014), the authors took a more applied approach to explore the possible applications of using ISC to detect user engagement. They used stimuli from popular culture in the shape of the pilot episode of the hit TV-series *The Walking dead* and commercials aired during the sporting event, Super bowl, usually the most-watched American television broadcast of the year. As a measure of user engagement the authors used Twitter activity, viewership and user ratings. The result was that the behaviour of thousands upon thousands of people, interacting with the stimuli in their everyday lives, could be significantly related to the ISC of study groups from 12 or 16 subjects. The ISC could even be related to a drop in activity during commercial breaks.

Lucas' group has since then further investigated the relationship between attention and ISC by manipulating attention further. They introduced audio-only stimuli as well as paradigms where the subjects counted backwards or had their viewing constrained by a fixation cross (Ki et al., 2016).

2.2.2 Correlated component analysis

Here follows a brief explanation of CorrCA. For a more thorough review of CorrCA and its extensions I recommend Parra (2018) or Parra et al. (2018).

CorrCA was introduced in Dmochowski et al. (2012) in order to transfer the concept of ISC in fMRI data to EEG analysis. Figure 2.1 attempts to illustrate the concept of the method, where the goal is to find a common spatial filter that transforms the multi-dimensional EEG into a single component, in a way such that the components from the EEG of each subject are maximally correlated.

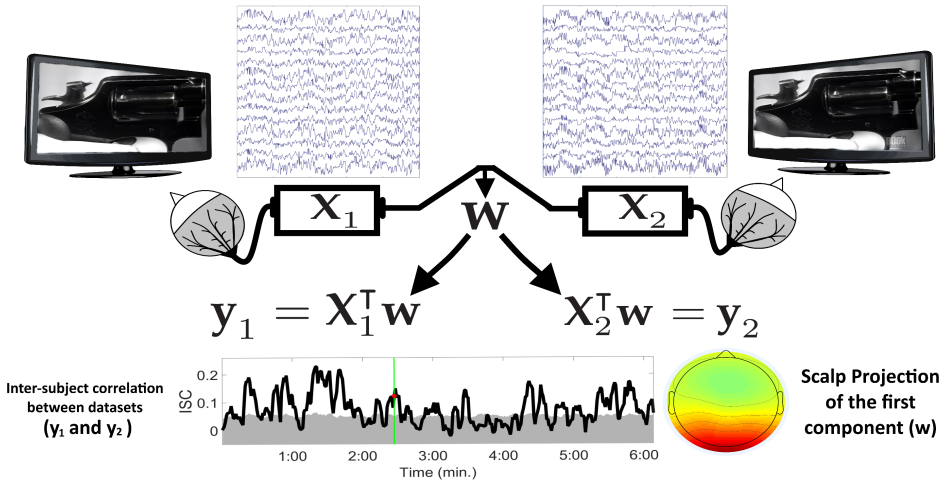


Figure 2.1: Conceptual illustration of using CorrCA for estimating inter-subject correlation. Two subjects see the same film, in this case a short film by Alfred Hitchcock, while they have EEG recorded ($\mathbf{X}_1$ and $\mathbf{X}_2$). The EEG is usually synchronised offline, but could also be recorded from subjects watching the film together. CorrCA then finds a spatial filter, $\mathbf{w}$, in order to maximise the correlation between the resulting components, $\mathbf{y}_1$ and $\mathbf{y}_2$. The ISC is calculated as the average correlation between all subject pairs. Here the ISC is calculated for 5-second windows of EEG with a 80 % overlap to ensure a temporal resolution of one second to obtain an ISC time curve as shown at the bottom. The red dot on the ISC curve signifies the timing of the scene shown on the screens as well as the five seconds of EEG shown for each subject.

CorrCA is derived as a special case of canonical correlation analysis (CCA, Hotelling, 1936), so we will start by looking at how this method works, and then how CorrCA differentiates itself from it. CCA is conceptually related to principal component analysis, using eigenvalue decompositions of covariance matrices. CCA is usually used to model correlated sources in two different modalities, and therefore estimates spatial filters, $\mathbf{w}_1$ and $\mathbf{w}_2$, for the input dimensions of each dataset.

If we denote the EEG from subject 1 and 2 as $\mathbf{X}_1$ and $\mathbf{X}_2$ the CCA components are defined as $\mathbf{y}_1 = \mathbf{X}_1^\intercal \mathbf{w}_1$ and $\mathbf{y}_2 = \mathbf{X}_2^\intercal \mathbf{w}_2$. These filters are found such that they maximise the correlation between the components:

$$\rho = \underset{\mathbf{w}_1,\mathbf{w}_2}{\arg\max} \frac{\mathbf{y}_1^\intercal \mathbf{y}_2}{\|\mathbf{y}_1\|\|\mathbf{y}_2\|} \ . \tag{2.1}$$

Introducing the sample covariance matrix, $\mathbf{R}_{ij} = \frac{1}{N}\mathbf{X}^i\mathbf{X}_j^\intercal$, (2.1) can be expressed as:

$$\rho = \underset{\mathbf{w}_1,\mathbf{w}_2}{\arg\max} \frac{\mathbf{w}_1^\intercal \mathbf{R}_{12}\mathbf{w}_2}{\sqrt{\mathbf{w}_1^\intercal \mathbf{R}_{11}\mathbf{w}_1}\sqrt{\mathbf{w}_2^\intercal \mathbf{R}_{22}\mathbf{w}_2}} \ , \tag{2.2}$$

that can be solved with two eigenvalue decompositions (Hardoon et al., 2004). CCA finds multiple components (dependent on the input dimension of $\mathbf{X}_1$ and $\mathbf{X}_2$), and constrains the estimation of the weights with the condition that the components are mutually uncorrelated (Klami, 2013).

The concept of CorrCA can be seen as a simplification of CCA, by only having one shared spatial filter, $\mathbf{w}$. This assumption makes the method unusable for many applications of CCA, where different types of datasets are compared. However, it works when comparing EEG datasets where the electrodes are in general assumed to be in similar positions.

Restricting the model to a common spatial filter has two benefits. It means that there are more degrees of freedom to estimate the filters and it actually makes it easier to calculate ISC for more than two subjects simultaneously. For CorrCA the equation corresponding to (2.2) is changed to:

$$\rho = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{R}_{12} \mathbf{w}}{\sqrt{\mathbf{w}^T \mathbf{R}_{11} \mathbf{w}} \sqrt{\mathbf{w}^T \mathbf{R}_{22} \mathbf{w}}} \ , \tag{2.3}$$

that can now be solved with a single eigenvalue decomposition:

$$\left(\mathbf{R}_{11} + \mathbf{R}_{22}\right)^{-1} \left(\mathbf{R}_{12} + \mathbf{R}_{21}\right) \mathbf{w} = 2 \cdot \frac{\sigma_{12}}{\sigma_{11}} \mathbf{w} \ , \tag{2.4}$$

where $\sigma_{ij} = \mathbf{w}^T \mathbf{R}_{ij} \mathbf{w}$.

CorrCA is a relatively simple model machine learning-wise. In Kamronn et al. (2015)[1], we sought to make a more advanced model, Bayesian CorrCA (BCorrCA), that combined the attributes of both CorrCA and CCA. Like CCA, BCorrCA estimates a spatial filter for each dataset, but also estimates a common filter like in CorrCA. The model was created with the rationale that, though the electrode positions and general cortical structure can be assumed similar between subjects, there might be smaller differences, e.g. due to the fitting of the EEG cap or slight differences in cortical dipole positions. Using Bayesian inference BCorrCA seeks to estimate how similar each subject-filter is to the shared filter and, unlike standard CCA, allows for analysis of more than two datasets at a time. Also, being a Bayesian model enables BCorrCA to estimate the forward model directly, unlike CCA and CorrCa that estimates the backward model (see Haufe et al., 2014, for a discussion on the difference between forward and backward models).

For the analysis in Poulsen et al. (2017) we decided to use the original version of CorrCA, since its simplicity keeps the computational cost low, and for comparability with the results of the original study. The low computational cost is a factor that will be relevant for future real-time experiments.

2.3 EEG in the classroom

In our article we wanted to investigate whether it was possible to reproduce the results of Dmochowski et al. (2012), but using wireless low-cost equipment in a classroom environment. Additionally, we wanted to investigate the feasibility of simultaneously recording EEG from multiple subjects, making it possible to create a shared experience.

[1] This work was performed during our master thesis and finished prior to this PhD.


Figure 2.2: In the joint viewings nine subjects viewed films together, while we recorded EEG from each of them simultaneously. **(Left):** The nine subjects where placed in a line to induce a cinema-like experiences. **(Right):** The joint viewing subjects watched the films projected onto a screen. The picture shows the tablets, that recorded the EEG, resting on tables behind the subjects. Each subject had a designated tablet that the EEG was transmitted wirelessly to from their cap. Figure from A.

2.3.1 Bringing EEG out of the laboratory

In our study we played two film clips and a control video for the subjects. As in the original study we used *Bang! You're dead*, and we added a clip from the film *Sophie's Choice*.

We had four groups of subjects that watched the films under three different conditions. One group ($N = 12$) watched the films individually in an university office on a tablet with a 7" (17.8 cm) screen using earphones. The second group ($N = 12$) was recorded in the same environment as the first group, but with a different version of the films. In this version the scenes were scrambled in time, to induce a loss of narrative. The final two groups watched the films together in groups of nine on a screen in a classroom (see figure 2.2), with sound projected through loudspeakers.

Instead of research-grade EEG equipment, like the 64 channel Biosemi system used in the original study, we employed the Smartphone Brain Scanner (SBS) system. The SBS is a modified 14 channel system, based on the Emotiv EPOC headset. In this study the SBS was implemented on 2013 versions of the Asus Nexus 7 tablets.

Though the price of using a low-cost, portable system outside an electrically shielded room is a loss in signal quality and lower temporal resolution[2], the quality of the SBS had already been proved sufficient for basic neuro-feedback paradigms (Stopczynski et al., 2014a,b). In our study we sought to test whether the signal-quality was also usable for more complex cognitive paradigms.

One of the issues we encountered with the SBS, was ensuring synchronisation of the EEG with the stimuli. Since CorrCA correlates raw EEG without any temporal smoothing, even a difference of a couple of samples in the synchronisation between EEG datasets can have detrimental effects on the ISC. Among the 42 EEG datasets we

[2]The SBS system was based on the first generation of Emotiv EPOCs that had a sampling frequency of 128 Hz.

recorded, nine datasets had to be excluded due to instability of the wireless connection that made it impossible to synchronise the data across subjects. To avoid imbalance in the analyses due to unequal number of subjects in each group, we excluded five additional subjects (selected at random), ultimately ending up with seven subjects for each group with synchronisable EEG.

Another issue compared to the original study was that the SBS does not include an electrooculogram (EOG), which the original study used to regress out eye related artefacts. This, however, was not a significant drawback, as independent component analysis (ICA) is effective for removing eye related artefacts. The EOG electrodes might even be considered obtrusive and could break immersion in paradigms situated in the everyday scenarios. To remove eye related components in a semi-supervised manner we employed the Corrmap plug-in for EEGLAB, removing up to 3 of the 14 available independent components (Delorme and Makeig, 2004; Viola et al., 2009).

In our analysis we could reproduce a drop in ISC from the original to the scrambled version of the films, as well as a lower ISC for an uneventful baseline video of people descending an escalator. This showed that the the lower grade equipment was able to get similar results as those obtained in Dmochowski et al. (2012). However, more interestingly we tried to correlate the ISC curves we obtained with the ones from the original study, resulting in very significant correlations (between 0.51 to 0.61, depending on viewing group).

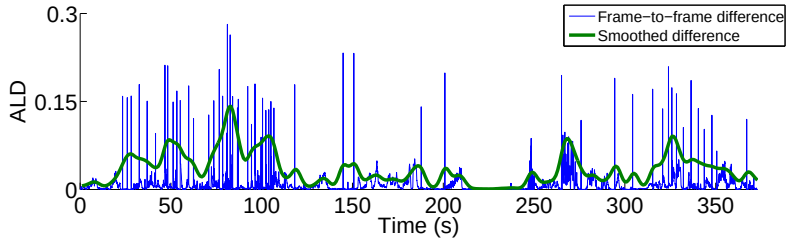
We interpreted these results as proving that neuroscientific methods can be moved into the classroom.

2.3.2 Investigating the neural origin of ISC in EEG

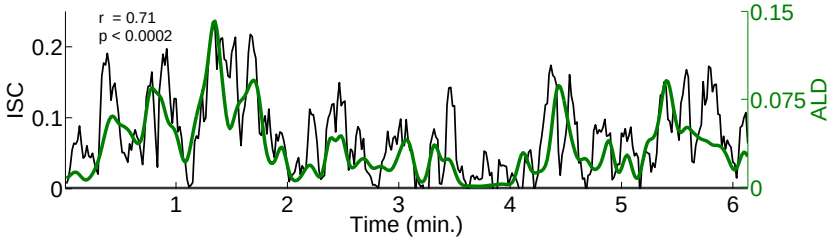
We were curious as to how a film could elicit neural responses that were so similar that the resulting ISC curves could be reproduced by lower grade equipment, in a less controlled setting, and by an independent research group on another continent. The previous studies indicated that there was a relation between ISC and attention or engagement, but the physiological link between the two was not clear to us.

There are many types of stimuli when watching a film, such as the visual, auditory or semantic content. One of the things we wanted to investigate was the relation to something as basic as the variations in the brightness of the film. One of the things we investigated was the frame-to-frame difference in pixel intensity of the videos. We averaged over all the pixels of a frame to obtain the univariate feature, which can be seen in figure 2.3(a). We could immediately see two things. First, the large peaks in the luminance feature, that corresponds to scene cuts, co-occured with increases in ISC. Second, it was hard to compare the feature with ISC due to differences in sampling frequency and the large changes in the luminance feature for frames during a scene and the ones following a scene cut.

In order to make it easier to compare the ISC with the changes in luminance we calculated the average luminance difference (ALD). The ALD is calculated by a non-linear down-sampling of the frame-to-frame difference to the sampling frequency of ISC (1 Hz) by selecting the maximum value for each 1-second interval in order to



(a) ALD with average frame-to-frame luminance difference.



(b) ALD with ISC.

Figure 2.3: ALD compared with the average frame-to-frame luminance difference of *Bang! You're Dead*, and with the ISC from subjects viewing the same film. **(a):** Comparison between the ALD and the average frame-to-frame luminance difference calculated from *Bang! You're Dead*. The ALD is calculated from the frame-to-frame difference in pixel intensity by sub-sampling it ($f_s = 1$ Hz) and smoothing it to match the 5 s window of ISC. Each large value of the frame-to-frame difference corresponds to a change in camera position (or scene cut). **(b):** The ISC of the first CorrCA component is temporally correlated with the changes to luminance of the film, as measured by ALD. Figures from A.

emphasise the large differences during scene cuts. Additionally, the ISC was calculated in 5-second windows (with 80 % overlap). In order to make the ALD relate to a similar window size, the down-sampled frame-to-frame difference was smoothed temporally by convolving it with a Gaussian kernel with a variance parameter of 2.5 s^2. This resulted in the ALD curve seen in figure 2.3.

Figure 2.3(b) shows that there is a very strong relationship between the luminance of the film viewed by the subjects, as measured by ALD, and the ISC calculated from their neural responses. On the surface this is a "good news/bad news" situation for the ISC as an indicator of attention. The good news is that this shows a link between ISC and a very well established electrophysiological phenomenon: The strength of visual evoked potentials (VEPs) are related to the strength of the stimulus. The bad news is that these are low-level responses, as opposed to a high-level process like attention.

Our interpretation of this result is that there is not a direct link between ISC and attention, as ISC seems to be driven by low-level responses. However, our results indicate that there is an indirect relationship. Subjects that were exposed to the same visual stimuli by watching the scrambled scenes exhibited lower ISC, which can be attributed to lower attention due to the loss of narrative. Using this reasoning the

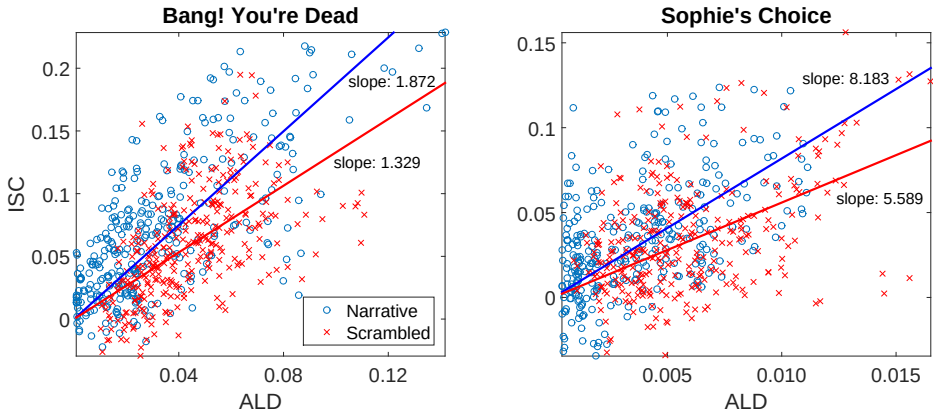


Figure 2.4: Relationship between the ISC and the changes in stimulus luminance for different conditions. Each point indicates a time point in the film with corresponding ISC and ALD value as seen for *Bang! You're Dead* in figure 2.3(b). It is evident that time points where the film excibited a higher degree of changes in luminance (high ALD) resulted in higher correlation of brain activity across subjects (high ISC). The indicated "slope" is calculated from a least squares fit with the lines passing through (0,0), and indicates the strength of ISC for a given ALD value. For both films there is a significant drop in the slope ($p < 0.01$: block permutation test with block size $B = 25$ s), from the original narrative (blue) compared to the scrambled version (red). Note that brightness of the scenes in *Sophie's Choice* is much lower than in *Bang! You're dead*, resulting in an ALD that is lower by almost a factor 10. Figure from A.

ISC can be seen as an indirect marker of attention, through a top-down modulation of the VEPs.

Various cases of attentional modulation of neural activity are well-established in the scientific literature. It has been shown that VEPs are modulated by spatial attention (Johannes et al., 1995) and that feature-specific attention enhances steady-state VEPs 25 s (Müller et al., 2006). Also, attentional modulation is not limited to either VEPs or EEG as the concept is prevalent in speech attention (Mesgarani and Chang, 2012; Mirkovic et al., 2015), and for ISC from fMRI (Lahnakoski et al., 2014).

2.3.3 Engagement to audio-visual stimuli modulates ISC

To further investigate whether ISC is modulated by subject attention to the stimulus, we looked to the "scrambled" condition. This condition was introduced in order to demonstrate that the correlation of neural responses across subjects was not simply the result of low-level stimulus features. The original and scrambled films exposed the subjects to identical audio-visual stimuli, so if the ISC was purely reflecting low-level responses to changes in luminance, the two conditions should elicit the same degree of ISC. However, in the original study a significant decrease in ISC was seen for the scrambled condition, presumably because the disrupted narrative coherence decreased the engagement of the subjects. As mentioned, we were also able to reproduce this effect in our experiments.

From figure 2.3(b) it is evident that peaks in ISC for *Bang! You're Dead* coincides with

the intervals containing many scene cuts. The relationship between scene cuts, ISC and attention could, however, also be due to more complex interactions, as fast-paced cutting is a well-known cinematographic tool used by Hitchcock to induce suspense and thereby increase the attention of the viewer (Bordwell, 2002). This interaction is supported by post-experiment questionnaires, where we asked subjects to describe the scenes that made the biggest impact on them. We quantified their answers by assigning each answer to one of eight general scene descriptions. The most frequently mentioned scene occurs around 2:25, where there is a peak in the ISC, and which was also confirmed by the suspense ratings presented in Naci et al. (2015).

In an effort to disentangle the attentional modulation of ISC, we looked at the linear relationship between ISC and the luminance fluctuations as quantified by ALD. Figure 2.4 shows the ALD and ISC for each time point in the two film clips from *Bang! You're Dead* and *Sophie's Choice*. For both films we could register a significant decrease in the ISC/ALD linear relationship. This drop can be interpreted as an attenuation (or lack of amplification) of VEPs due to lower subject attention to the films with scrambled scenes and disrupted narrative.

2.4 Current status of ISC as a marker of attention

Our study did not feature an actual educational scenario, but focused on the feasibility of reproducing the results of laboratory experiments in a classroom, and to record EEG from many subjects simultaneously. The experiments were conducted in late 2013, and simultaneous or subsequent studies have brought the EEG and attentional measures further into the classroom in educational paradigms.

Dikker et al. (2017) used consumer grade devices to record EEG from high school students during an entire semester of biology classes and related neural synchronisation to teaching styles and group interactions in the classroom. Another study introduced attention-related tasks in lectures to relate frequency-based neural markers to readiness and attention (Ko et al., 2017).

Lucas Parra's group has also been very productive in exploring the capabilities of using ISC in learning paradigms, and not just in a strict attentional context. For example Cohen and Parra (2016) showed that the strength of ISC is predictive of memory performance after 3 weeks, using ISC to track how in-synch a subject was with other subjects at any given time, and applying it as a marker of attention. This has been expanded to an online learning paradigm, where ISC correlated with performance in a subsequent test, even when subjects were unaware that they were going to be tested (Cohen et al., 2018).

To summarise, our study is one of several current studies which show that EEG can be used in a classroom. That it is possible to record EEG of sufficient quality and that ISC is a viable marker in tracking attention. One of the next questions is whether ISC by itself contains enough information to be used in an educational context, or if there is a possible gain from combining ISC with other EEG neural markers. Examples of relevant neural markers are the EEG complexity, as discussed in chapter 4, and the methods being developed for tracking attention to speech (e.g. Mesgarani and Chang,

2012; Wong et al., 2018). Or perhaps we should even combine ISC from EEG with other biological signals?

CHAPTER 3
Microstate analysis

This chapter describes how we created a user-friendly toolbox for microstate analysis with an emphasis on transparency of the underlying methods. The chapter also covers how we, during this process, sorted through the methods employed in the relatively old field of microstate analysis, and the methodological guide that resulted from it (Poulsen et al., 2018). This chapter does not contain any new analysis methods, though the toolbox did end up featuring an unpublished variational model by Frans Zdyb and an optimised iteration scheme for modified K-means. This chapter will therefore not include discussions on the interpretations of microstate analysis. The article is available to the reader in appendix B.

This work was conducted in collaboration with Andreas Pedroni and Nicolas Langer of Zürich university.

3.1 An introductory guide to microstate analysis

We originally intended this to be a small project, with the goal of enabling users of the EEGlab toolbox for Matlab (Delorme and Makeig, 2004) to run an experimental clustering method and compare it with modified K-means (Pascual-Marqui et al., 1995). However, it ended up being a collaboration across universities that resulted in a much more extensive and versatile toolbox and in a guide spanning 30 pages.

One reason for the expansion of the project was that our ambitions grew from a very synergistic collaboration. I came into the project with a general experience with machine learning and EEG analysis, but a naïve approach to the field of microstate analysis, which enabled me to wonder about standard practises that were otherwise taken for granted. Andreas and Nicolas had both the experience and methodological interest, so that we together could build a transparent toolbox that both catered to experienced users of microstate analysis and had a solid foundation in machine learning.

The second reason for the expansion of our accompanying guide, was a wish to make a general introduction to researchers that are new to the field, and who have earlier been

hesitant due to a lack of tools for running microstate analysis in Matlab, and also due to the somewhat opaque methodology. Our guide has therefore ended up containing an overview of the different components of microstate analysis: Clustering methods and their settings; measures of fit to select the number of clusters; backfitting microstate prototypes to EEG; temporal smoothing of microstate labels; and microstate statistics.

Finally, while going through the microstate literature, that stretches more than four decades, I encountered several inconsistencies, such as unpublished changes to standard methods, citations of the wrong articles or even the wrong methods. We used the guide as a means to address some of these inconsistencies.

3.1.1 The concept of microstates

Microstate analysis is a way to represent EEG in a compact manner based on the topography of each EEG sample. Microstate analysis was founded by Dietrich Lehmann and colleagues (see e.g. Lehmann et al., 1987), by observing topographies that remained stable for 80 - 120 ms before rapidly transitioning to a different topography. These periods of quasi-stable EEG topography have been called functional microstates, with each spatial configuration representing a microstate class. Microstates are suggested to reflect global functional states in the brain, with their EEG topographies being the product of different configurations of neuronal generators (Khanna et al., 2015; Michel, 2009; Lehmann et al., 1998). Microstates have been found to be related to resting state networks from simultaneously recorded fMRI (see e.g. Van De Ville et al., 2010; Yuan et al., 2012; Britz et al., 2010).

3.1.2 Clustering methods in microstate analysis

The common concept of the different topographical clustering methods employed in microstate analysis, is the division of recorded EEG samples into microstate clusters, such that EEG samples within the same cluster have as similar topographies as possible. A prototypical topographical map is then calculated for each cluster, based on all the EEG samples assigned to it. It is then assumed that these prototypes then represents the spatial distribution and neural processes of all EEG samples assigned to them.

In our guide we review the three most common clustering methods used in microstate analysis, namely *K-means*, *modified K-means* and *(Topographic) Atomize and Agglomerate Hierarchical Clustering - (T)AAHC*.

As mentioned earlier, I encountered inconsistencies or methodological issues, while reviewing the literature. Two of the areas of confusion, which we address in our manuscript, are covered in sections 3.1.3 and 3.1.4.

3.1.3 Selecting the number of microstates

In microstate analysis, and in clustering in general, there is a challenge of selecting the "correct" number of clusters to explain the EEG. In microstate analysis, this is often

addressed by testing a range of different numbers of clusters, and then measuring how well the resulting prototypes explain the data by using measures of fit.

Among these measures of fit, the Krzanowski-Lai (KL) criterion is quite popular. It was introduced to select the number of clusters based on their dispersion measures, W. W is calculated as the sum of squares between all possible pairs of members belonging to the same cluster, and is supposed to decrease monotonously with increasing numbers of clusters. KL seeks to find the "elbow" in the W curve, that signifies that adding another cluster will give a relatively low increase in fit.

However, microstate analysis is sometimes conducted using polarity-invariant methods[1], and we argue that for these methods the W, and thereby the KL criterion, is an ill choice as a measure of fit. This can be explained using a simplified example, with a microstate cluster that consists of a prototype, $\mathbf{a}$, and two members, $\{\mathbf{x}_1, \mathbf{x}_2\}$. Assuming an ideal situation where the members are equal to their prototype, $\mathbf{x}_1 = \mathbf{a} = \mathbf{x}_2$, the dispersion will then be:

$$W = ||\mathbf{x}_1 - \mathbf{x}_2||^2 = ||\mathbf{a} - \mathbf{a}||^2 = 0. \tag{3.1}$$

For a polarity-invariant method an, in principle, equally ideal situation would allow one of the members to have opposite polarity with respect to its prototype; $\mathbf{x}_1 = -\mathbf{a}$. However, for this case the dispersion does not reflect a good fit:

$$W = ||\mathbf{x}_1 - \mathbf{x}_2||^2 = ||-\mathbf{a} - \mathbf{a}||^2 = 4||\mathbf{a}||^2. \tag{3.2}$$

Contrary to standard practises in microstate analysis, I would therefore not recommend using the KL criterion for selecting the number of clusters for polarity-invariant methods. At least not using the standard dispersion measure. The dispersion measure could easily be modified to be polarity-invariant, e.g. by multiplying cluster members by the sign of their correlation with their prototype, which would make it suitable.

Another area, where the field of microstate analysis could benefit from an update, is with respect to cross-validation. Even though one of the measures of fit is called the cross-validation criterion, cross-validation, defined as dividing the data into a test and training set, is not common in microstate analysis. Instead, in microstate analysis one typically uses four common measures of fit to estimate how well the EEG samples fit the microstates they are assigned to. This could also be called training errors, and the question is whether the improvement of computational power and the current surge in machine learning, means that we are ready to introduce cross-validation using test errors to microstate analysis?

3.1.4 Confusion regarding the (T)AAHC methods

AAHC is a modification of agglomerative hierarchical clustering (AHC) in the traditional sense, where the main difference is how clusters are removed and agglomerated. Instead of merging the two most similar clusters, as in standard AHC (Rokach and Maimon, 2005), AAHC finds the "worst" cluster, which it then disbands (atomises)

[1] Modified K-means, AAHC and TAAHC are all polarity-invariant methods.

and assigns its members to the cluster prototypes they are individually most similar to. AAHC decides which cluster is the worst by measuring its global explained variance (GEV), which is one of the microstate-specific measures of fit mentioned in 3.1.3. The GEV is calculated as;

$$\mathrm{GEV}_n = \frac{(\mathrm{Corr}(\mathbf{x}_n, \mathbf{a}_{l_n}) \cdot \mathrm{GFP}_n)^2}{\sum_{n'}^{N} \mathrm{GFP}_{n'}^2}, \tag{3.3}$$

where GFP_n is the global field power, which for each EEG sample is calculated as the standard deviation across all electrodes. The GEV can be seen as the squared correlation between an EEG sample and its microstate prototype weighted by that EEG sample's fraction of the squared GFP of all EEG samples:

$$\mathrm{GEV}_n = \mathrm{Corr}(\mathbf{x}_n, \mathbf{a}_{l_n})^2 \cdot \frac{\mathrm{GFP}_n^2}{\sum_{n'}^{N} \mathrm{GFP}_{n'}^2}. \tag{3.4}$$

To calculate the GEV for a specific cluster, you sum the GEV of all of its members. The higher the GEV the better the cluster.

All of this is quite nicely explained in Murray et al. (2008). However, subsequent to the publication of that article, the TAAHC was introduced. From a conceptual point of view, TAAHC only changes the measure of fit, that is used to find the "worst" cluster, from GEV to the summed correlations between a cluster prototype and its members;

$$\mathrm{CorrSum}(k) = \sum_{n}^{N} \mathrm{Corr}(\mathbf{a}_k, \mathbf{x}_n) = \sum_{n}^{N} \frac{|\mathbf{x}_n \cdot \mathbf{a}_k|}{||\mathbf{x}_n|| \cdot ||\mathbf{a}_k||}, \quad \text{for} \;\; l_n = k, \tag{3.5}$$

assuming average referenced EEG. This has only been described by the authors in the help pages for the stand-alone program, Cartool, by Denis Brunet. To my knowledge the first published explanation of the methodology behind TAAHC came in Khanna et al. (2014).

An issue for the TAAHC method is that neither of these two sources mentions, that by changing the measure of fit for the clusters, the TAAHC in principle becomes stochastic, unlike the deterministic AAHC. The source of the stochasticity lies in the initialisation of the agglomerative clustering methods, where each EEG sample has their own cluster. Therefore, the correlation between the cluster prototype and its single member will be exactly one for all clusters, and the first "worst" cluster would have to be chosen at random.

The reason why the stochasticity is only in principle, is that Cartool features an intermediate initialisation step, where all clusters start out with two members. The EEG samples are paired based on their correlation, and using this extra step the TAAHC becomes deterministic again. The issue, however, is that this step, to my knowledge, has not been published anywhere. I only became aware of it, because Denis Brunet was gratuitous with his time and explained it to us, when we contacted him after we became aware of the inconsistency while working on the guide and toolbox.

This, unfortunately, is not the only confusion surrounding the TAAHC method, as there seem to be some confusion in the field over which articles to reference for the method: Tibshirani and Walther (2005) is often referenced when the (T)AAHC methods are used, even though this article actually does not explain how agglomerative hierarchical clustering works. Also the (T)AAHC methods are also sometimes erroneously referenced to Pascual-Marqui et al. (1995), which pre-dates the methods and instead introduces the modified K-means algorithm, that has an entirely different approach to clustering. I also have a suspicion that some of the implementations of TAAHC, that are kept local at the different laboratories are either missing the determinism-ensuring initialisation step, or unknowingly use AAHC instead of TAAHC.

It is issues like these that motivated us to create an extensively documented toolbox for Matlab, so the methodology is transparent in the sense that it is easy to check the underlying code. It was also one of the reasons why our open-access guide to microstate methodology achieved its current considerable length.

3.2 Open-source toolbox for microstate analysis

Our motivation for creating a toolbox was to add transparency to the methods employed in the field of microstate analysis. With our guide we wanted to add clarity of the theoretical aspect of the methodology, and with our toolbox we wanted to ensure transparency in practical aspect of how the methodology was implemented. We wanted to achieve this transparency by ensuring that the code of our toolbox is open to scrutiny. Finally we wanted the toolbox to be an easy way to get started for newcomers, as well as increase the customisability and ability to run batch processes for advanced users of microstate analysis.

Though there are several programming languages, and accompanying neuroscientific toolboxes, that could fulfil these requirements, we chose to use Matlab and make an extension for the EEGlab toolbox. EEGlab has the benefit of being an established platform for EEG analysis and has both a user-friendly graphical user interface (GUI) as well as the ability to create scripts for batch processing of several datasets. Additionally, the `eegh` function makes it easy to make the jump from using the GUI to writing batch scripts, by printing the calls required to make the analysis just performed with the GUI.

We have made an emphasis on thoroughly documenting the code of the toolbox, both to make it easier to understand how the methods are implemented, and to be able to customise the toolbox e.g. by adding new clustering methods.

The toolbox is able to run microstate analysis on both ERP and spontaneous (e.g. resting state) EEG, using the following clustering algorithms:

- K-means.
- Modified K-means.
- Atomize and Agglomerate Hierarchical Clustering (AAHC).

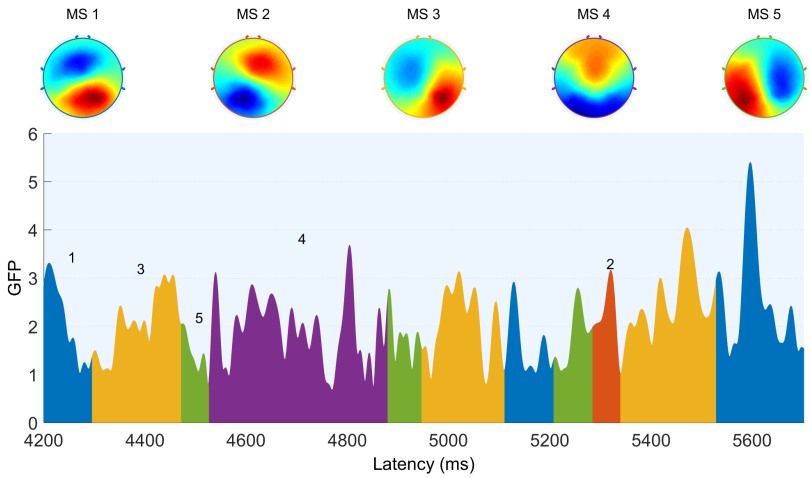


Figure 3.1: Illustrative figure of microstate segmentation for a 1,500 ms period of resting state activity. The EEG is represented by its GFP and the colours denote the active microstates for each time point. Prior to submitting the EEG to microstate analysis, the EEG was filtered with a high-pass filter of 1 Hz and a low-pass filter of 30 Hz using the standard settings of FIR filter of EEGLAB (v14.0.0). The figure is from B, and I refer to this article for more details regarding the steps in the microstate analysis.

- Topographic Atomize and Agglomerate Hierarchical Clustering (TAAHC).

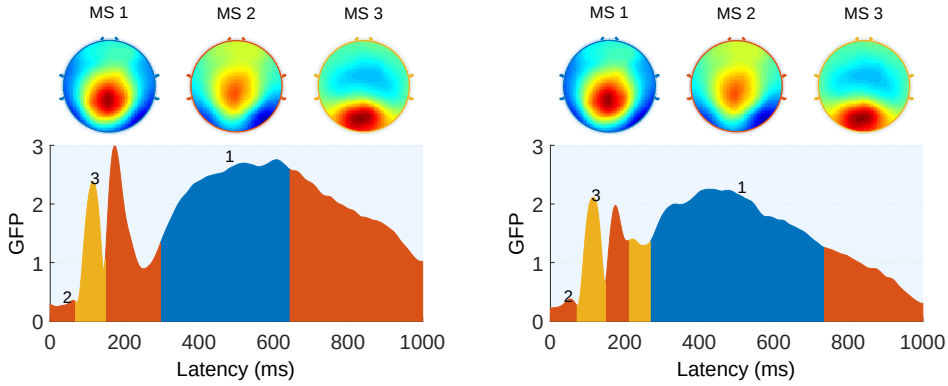
These four clustering methods are most commonly used in microstate analysis. Even though the motivation for the guide and toolbox is not to add new methods to the field, the toolbox did end up including two new additions to clustering microstates: An unpublished variational model by Frans Zdyb and an optimisation of iteration scheme in the modified K-means. In our preliminary experiments, the optimised modified K-means offered similar or slightly better performance[2] and a computational speed-up by a factor between 4 and 5 depending on the dimensionality of the dataset.

In addition to containing functions to perform statistics on the acquired microstates, the toolbox also contains the functionality to visualise EEG, as represented by its estimated microstates. Figures 3.1 and 3.2 show the microstate prototypes, GFP, and the most active microstates for each EEG sample.

Figure 3.1 shows an example of visualising spontaneous EEG, using the datasets of the tutorial in B, which consisted of resting state EEG from 4 subjects. The EEG originated from the data sharing paper by Langer et al. (2017).

An example of how to visualise and compare different conditions in an ERP analysis can be seen in figure 3.2. The data stems from Wakeman and Henson (2015) and consists of EEG from 16 subjects watching pictures of either famous faces or scrambled faces. For this example, the EEG was represented using few microstates to focus on

[2]Measured using the scaled mean square error between the noise free signal and the reconstructed signal, for simulated data. See B for more information.



(a) Famous faces.

(b) Scrambled faces.

Figure 3.2: ERP analysis of EEG from subjects watching pictures of either famous faces or spatially scrambled faces. Prior to submitting the EEG to microstate analysis, the EEG was filtered with a high-pass filter of 0.5 Hz and a low-pass filter of 30 Hz using the standard settings of FIR filter of EEGLAB (v14.0.0). The following microstate analysis steps were performed using the microstate toolbox: A grand average ERP was calculated for each condition. The ERPs from the two conditions were then concatenated and submitted to K-means using three clusters, to estimate the prototypes visualised above the microstate segmentations. In order to assign each EEG sample to a microstate, the prototypes where backfitted to each condition individually and smoothed using the *reject small segments* method.

differences in the N170 peak, which is captured by microstate 2. It can be seen that there is a clear difference in GFP amplitude between the two conditions around 170 ms, and that the prototype of this microstate can be seen to have a topography that is similar to the condition contrast presented in Ashburner et al. (2014, chapter 42) for 155 ms.

Both datasets were preprocessed using the Automagic[3] tool for artefact correction (with standard settings). See the figure captions for further information on the preprocessing of the data.

In summary, we have contributed to the field of microstate analysis with an open access guide and toolbox. We have created these with emphasis on transparency and thoroughness in the methodology, and we hope they will help make it easier for newcomers to get started in this field of EEG analysis.

[3] Available on its Github at: https://github.com/amirrezaw/automagic

CHAPTER 4
EEG complexity as a neural marker of cognitive processing

This chapter covers my work with EEG complexity. Unlike the other chapters, this work spans more than one study. The chapter starts with an introduction to the concept of calculating complexity of EEG and then covers my contribution to a study on semantic processing during sleep (Andrillon et al., 2016). Finally, my work on using EEG complexity to track user performance and difficulty in a helicopter simulator is covered. This work has not yet been published, but a draft manuscript and conference abstract describing our results has been added to the appendix. Because this study has not yet been published, I will present these results in greater detail. The manuscripts are available to the reader in appendices C, D and E.

4.1 EEG complexity

"You should call it entropy (...) nobody knows what entropy really is, so in a debate you will always have the advantage."
– John von Neumann (1903 - 1957)

Complexity, when used in neuroscience, can be both an exotic and abstruse concept. The fact that Lempel Ziv complexity can also be used as an estimator of entropy (see e.g. Jensen et al., 2010), does not make the concept less esoteric, as the quote above also indicates. However, the underlying calculations in EEG complexity are in principle quite straightforward, and I hope that this section can help demystify the measure and give some intuition on how it works.

There is, however, a clear distinction between how a method works and what the resulting measures mean for the underlying neurophysiology. Take for example the Fourier transform, that is often employed for frequency analysis. A simplified explanation of how the Fourier transform works, is that it seeks to explain a time series as a sum of sinusoids with different frequencies. Thereby it is possible to estimate

which frequencies dominate in this time series. It is much more difficult to give a definite answer to what changes in the frequency spectrum of EEG means. There exist many results from countless studies using frequency analysis for EEG but, to my knowledge, there is no consensus for a unifying theory that explains what frequency means with respect to the underlying neurophysiology. And this in spite of the fact that frequency analysis of EEG is almost as old as EEG itself (Hari and Puce, 2017).

It is the same situation for EEG complexity. I can explain how the method works and show how EEG complexity correlates with different mental states or levels of conscious processing. But I am not able to say with certainty, what EEG complexity means, as we lack both more results and a general consensus on this topic. However, I will give my personal intuitions on the matter when summarising the results.

4.1.1 Compressing EEG

EEG is an intricate multi-dimensional measure, and there are many approaches to analyse EEG or compare it with the stimuli that subjects are interacting with. In order to be able to interpret EEG it is usually necessary to transform the EEG into a more manageable measure, preferably a scalar value which e.g. can then be correlated with a given stimulus or contrast of conditions.

One such measure is EEG complexity, which gives an estimate of how compressible the EEG is. This compressibility can be related to entropy, or to the degree of predictability of the EEG. In general, a lowering of predictability of EEG has been attributed to higher states of awareness or conscious processing (see section 4.1.3).

To calculate the complexity, the Lempel-Ziv algorithm is used. Since this algorithm performs better with coarse or binary data, the EEG needs to be transformed into a binary string. There are different approaches to this, which will be covered in section 4.1.4, but first I will briefly explain the main steps.

Figure 4.3A illustrates the main steps of the pipeline to transform EEG into a binary string used in the studies presented in this chapter: A short segment of the EEG is selected and binarised based on a threshold calculated for each channel. This binary matrix is then "stretched" into a string by concatenating the observations one-by-one. Finally, the binary string is given to the Lempel-Ziv algorithm, which estimates the complexity of the string.

In the following section I will give an introduction to the concept of calculating complexity of EEG and how this method has been applied in neuroscience.

4.1.2 Lempel-Ziv complexity

The Lempel-Ziv algorithm can be used to calculate complexity of a signal by estimating how much it can be compressed without loosing any information (Lempel and Ziv, 1976). This is a well-known algorithm that is also used in known file lossless compression programs such as WinZip.

The algorithms examines a time series for repeated patterns, and can be said to estimate how predictable a time series is, based on past observations. It works by

seeing how few sub-strings of data that can be used to explain the entire time series. To do this it creates a "codebook" of all the sub-strings necessary, and the size of the codebook is then defined as the complexity for the time series.

For the results presented in this chapter I have used the implementation employed in Schartner et al. (2015)[1]. I denote complexity calculated from this implementation as LZc.

The algorithm starts at the first element in the time series, making sub-strings of the elements in the order they appear. Once a sub-string has been created, that does not exist in the codebook, the sub-string is added to the codebook, and a new sub-string is created starting at the element it had reached in the time series. In this implementation, the algorithm keeps the previous element in memory, and includes it when starting on a new entry for the codebook.

As an illustration of how the algorithm works, (4.1) shows how the codebook and complexity for a short binary string:

$$\begin{aligned} \text{string} &= 0011110000111000011110011000111 10. \\ \text{Codebook} &= \{\ 0 \mid 00 \mid 01 \mid 11 \mid 111 \mid 10 \mid 000 \mid 001 \mid 1110 \mid 0000 \mid 011 \mid \\ &\qquad 11100 \mid 0110 \mid 0001 \mid 1111\ \}. \\ \text{LZc}(\text{string}) &= \text{length}(\text{Codebook}) = 15. \end{aligned} \tag{4.1}$$

To illustrate the difference in complexity between predictable and random strings, imagine three strings with 1000 elements each:

$$\begin{aligned} s_1 &= 0000000000000000000\ ... \\ s_2 &= 10101010101010101010\ ... \\ s_3 &= 01111000010110011110\ ... \end{aligned}$$

Here, the constant string, s_1, has an estimated complexity of: $\text{LZc}(s_1) = 45$; s_2, which alternates between 1 and 0 has a complexity of $\text{LZc}(s_2) = 63$; and the random s_3 is estimated to have a much higher complexity of $\text{LZc}(s_3) = 195$. See Jing Hu et al. (2006) for methodological review regarding complexity for strings of finite length.

One thing that is noticeable between the two examples is that the estimated complexity is very dependent on the length of the string. To be able to compare different lengths of strings, and ease interpretation, complexity is often sought normalised to values between 0 (fully compressible, low degree of randomness) and 1 (minimally compressible, high degree of randomness).

This normalisation can be done using Shannon entropy, which should make strictly random sequences asymptotically equivalent to 1 for sequences of infinite length (see e.g. Casali et al., 2013). However, in my experience this asymptotic behaviour requires

[1] The authors have shared their python code for the algorithm, which I have translated into a Matlab script.

more samples, than available in the short windows of EEG, which I use in my analyses. I have therefore employed the normalisation scheme of Schartner et al. (2015) in which the LZc is normalised by the complexity of the same string, where the elements have been shuffled randomly.

4.1.3 Complexity applied in neuroscience

Recent research has shown that the complexity of EEG signals is correlated with the levels of consciousness in comatose patients (Casali et al., 2013) as well as healthy subjects under anaesthesia or during sleep (Schartner et al., 2015; Andrillon et al., 2016; Schartner et al., 2017).

However, the Lempel-Ziv algorithm is more than four decades old, and many studies have used it to calculate complexity of EEG signals earlier. Some of the findings were that EEG complexity was influenced by anaesthesia or neurological diseases (see e.g. Zhang et al., 2001; Abasolo et al., 2006; Li et al., 2008). However, these studies calculated the complexity on individual electrodes, and therefore left out any spatial information in the EEG.

By including the spatial information, Casali et al. (2013) could interpret the EEG complexity in terms of loss of integration (reduced interaction among cortical areas) and loss of differentiation (many interacting areas all react to the perturbation in a stereotypic way). The authors also investigated the effect of transposing the binary EEG matrix before calculating complexity and found a very strong correlation with the complexity of the original matrices. They interpreted this correlation to demonstrate that the complexity *"is sensitive to both the spatial and temporal dimensions (...)"*.

In a study using intracranial EEG, the authors investigated the effect of including spatial information by calculating LZsum, the sum of complexity of each individual channel (Schartner et al., 2017). LZsum was in principle calculated on the same binary matrix as LZc, but excluded the spatial information. They found that LZsum was still able to contrast sleep stages, which is unsurprising as sleep stages in large part are defined by their temporal waveforms (see e.g. Iber et al., 2007). However, they also reported that the LZc was slightly better than the LZsum at separating sleep stages, based on the same EEG. This indicates that it is beneficial to calculate EEG complexity in a manner that includes spatial information.

To summarize, while there are not many methodological investigations into the spatial component in EEG complexity, the few available studies indicate that there is a benefit from calculating the complexity on EEG channels together.

4.1.4 Approaches to binarising EEG

There exist different experimental paradigms with respect to recording conditions and what type of EEG is passed to the Lempel-Ziv algorithm. In Casali et al. (2013), the authors use complexity in their perturbational complexity index (PCI). Briefly, they stimulate subjects several times with transcranial magnetic stimulation (TMS), averaging over the events to get an ERP. Using source modelling and non-parametric statistics, they binarised the ERP matrix before calculating its complexity.

Schartner et al. (2015) was inspired by this study, but diverged by calculating complexity on spontaneous EEG in 10-second segments. They also used a more common approach to binarise the EEG. For each segment, the mean of the EEG signal was subtracted from each channel and linearly de-trended. The Hilbert transform was then applied, and samples above the channel-average were set 1, and 0 otherwise.

There are obvious possible modifications to this binarisation-scheme, and their influence on the resulting complexity is an interesting avenue of research, that I unfortunately were unable to investigate due to time-constraints.

If LZc is to have a more widespread application in the field of neuroscience, more methodological investigates of what LZc captures are needed. For example, how is the complexity influenced by the sampling frequency or the choice of binarisation threshold (or binarisation method in general)?

Some of these questions have already been investigated for complexity of single-dimensional time series of biomedical origin (see e.g. Aboy et al., 2006; Jing Hu et al., 2006), and King et al. (2013) showed how the sampling frequency determined the frequency reflected by their weighted symbolic mutual information measure. However, it has not been investigated how these questions are affected for complexity methods including spatial information. For example, I hypothesise that there might exist a trade-off between the number of electrodes to include, and the length of the temporal window and the sampling frequency.

In the studies presented in this chapter, I have used a paradigm that is close to the one presented in Schartner et al. (2015), with the important modification that we investigated shorter segments of EEG. With a window length of 500 or 1,500 ms slid 50 or 100 ms per LZc sample, we were able to investigate the temporal changes in complexity with respect to the experimental paradigms.

In both studies we pre-processed the EEG before binarisation. First, the raw EEG was filtered using a 85 Hz lowpass and a 50 Hz bandstop FIR filter to ensure linear phase and avoid phase distortion in the EEG. Finally, we used surface Laplacian, to reduce the influence of volume conduction (BCILAB plug-in for EEGLAB, Delorme and Makeig, 2004; Kothe and Makeig, 2013).

4.2 Complexity in sleep

Sleep is an interesting phenomenon. Even though researchers are continuously learning more about how and why we sleep, there is still much we are unsure of. In this project I was lucky enough to be part of a brilliant[2] study tackling the question: *Does unresponsiveness to external stimuli during sleep mean that sleepers are isolated from sensory input from their environment?*

In order to examine how connected the sleeping brain is from its environment, Kouider et al. (2014) devised a paradigm to study the ability of the sleeping brain to conduct semantic processing.

[2] I was not part of the planning, nor execution of the experimental paradigm, so I can say this without being smug.

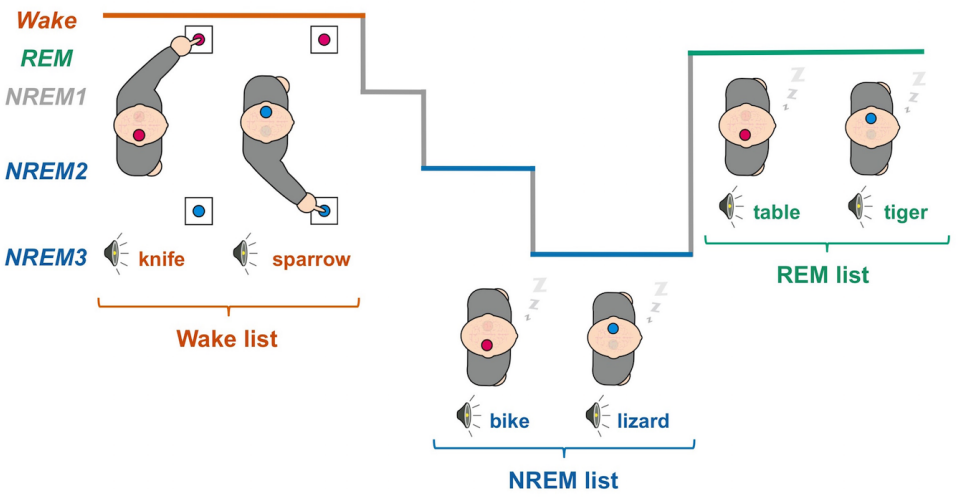


Figure 4.1: Illustration of experimental protocol. Subjects were instructed to classify words through left or right hand responses according the semantic category of the word (animal or object). Different lists of words were played depending on which stage of sleep they were currently in. One list for NREM sleep (NREM2 and NREM3, blue), one for REM sleep (green), and wake list (red) was played otherwise. Using lateralisation of motor-related brain activity (EEG), it was possible to examine whether the subjects were able to process the semantics of the words up to the level, where their cortical area, of the corresponding hand, was activated. Figure from C.

In Andrillon et al. (2016) this study was expanded to a full night experiment, to how the brain's responsiveness to its environment changes through the different stages of sleep. Specifically, the sleep was divided into rapid eye movement (REM), light and deep non-REM (NREM) and wakefulness.

As my role in this study was limited to the complexity analysis of the recorded EEG, I will focus on this aspect of the results, and in general keep this section brief. I refer to the article for a full description of all the results.

4.2.1 Experimental setup

The basic concept of the experiment is visualised in figure 4.1. Different words were played for the subject, who was instructed to make semantic decision task by indicating the category of each word through a response with the right or left hand. The two categories of words were objects and animals. This lasted for a full night of sleep, while the subjects lay in bed and had their EEG recorded.

The subject was awake when the experiment started, and was encouraged to fall a sleep during the experiment. The subject was also instructed to continue to make a button-response if they woke during the night.

While this sounds simple, the execution was not so. To be sure that the responses

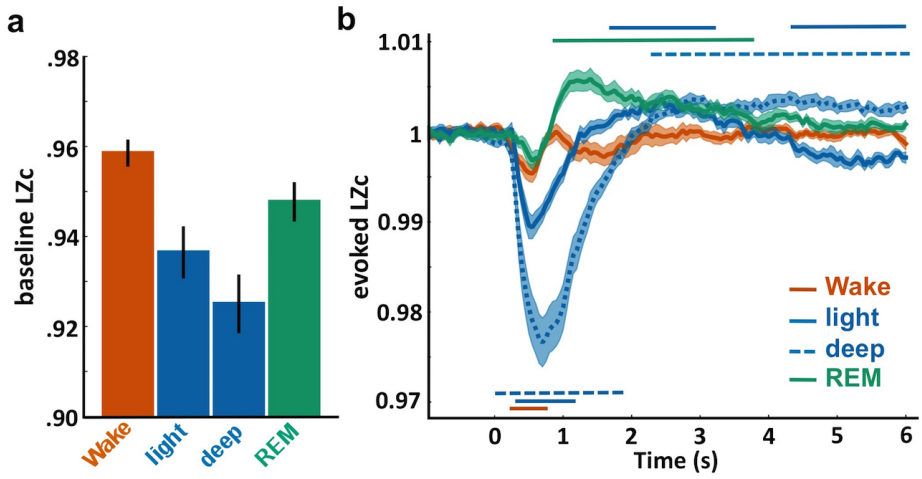


Figure 4.2: LZc across sleep stages, with error bars indicating the SEM over subjects. **a:** Baseline complexity varies with sleep stages. The baseline was calculated using prestimulus activity (-1.5 to 0 s) averaged across trials. **b:** EEG complexity is modulated by stimulus presentation. Evoked complexity was calculated by locking LZc to stimulus onset and expressed as a ratio of the baseline level. Horizontal lines denote significant deflections using cluster permutation at a $p_{\text{cluster}} < 0.05$ level. Figure from C.

were due to semantic processing and not simply stimulus-response associations learned while awake, three lists of words were used during the night. Words from each of the three lists were only played during a specific sleep stage.

This meant that Thomas (the lead author of the article) had to monitor the EEG of the sleeping subjects to conduct sleep-scoring in real-time, so he could switch between the word lists being played.

Twenty-three subjects participated in the study ensuring many sleepless nights for Thomas. Of these five participants were discarded from our analyses due to either technical issues or because subjects experienced difficulties falling asleep.

4.2.2 Results

Semantic processing of the words played during sleep was examined using an EEG marker of motor preparation, the lateralised readiness potential (LRP). The LRP is calculated by contrasting EEG from the motor area of the hand assigned to the word category with the EEG related to the other hand. I contributed to the study with an additional analysis based on the EEG complexity. The main results from the LZc are summarised in figure 4.2.

The LZc calculated from the baseline EEG prior to stimuli allowed to unambiguously separate the different states of sleep (and vigilance). The statistical significance of the separation was shown with a one-way ANOVA ($p < 2 \cdot 10^{-5}$), N = 18 subjects) and paired t-tests between states ($p < 0.005$, uncorrected for multiple comparisons).

Furthermore, I calculated the evoked LZc by using short windows of 500 ms and high overlap, giving a temporal resolution of 50 ms. From the resulting LZc time curves, we could calculate the event-related complexity by averaging across trials in the same manner as for ERPs. A cluster permutation test showed that stimuli robustly modulated the complexity of the EEG with an initial decrease after stimulus onset ($p_{\text{cluster}} < 0.05$), except for REM sleep). This initial decrease was followed by an increase in complexity in light NREM, deep NREM, and REM sleep ($p_{\text{cluster}} < 0.05$).

By correlating the baseline LZc with the LRP magnitude, we could show that there was a relationship between the EEG complexity prior to stimuli and the strength of the motor-related response. Computed across the entire night, we found a significant correlation for the wake, light NREM, and REM sleep trials ($p < 0.005$ for all). However, for the REM stage the relationship was inverted compared to wake and light NREM.

4.3 Tracking user performance and task difficulty in a helicopter simulator

In this study we investigated whether LZc can be used in fully aware, healthy people as an index of how focused they are on a given task.

Twenty subjects (hereof 10 were female) were recruited for an experiment, where they had to use a helicopter simulator to navigate through courses with varying difficulty with the goal of flying through circles. While the subjects interacted with the simulator, we recorded their EEG in order to investigate whether their neural activity reflected their performance of navigating the helicopter, as well as the varying difficulty of the simulator.

Figure 4.3 illustrates the concepts of the experimental setup and complexity analysis of the EEG.

4.3.1 Experimental setup

The experiment was constructed in the following way: A *trial* is defined as the time spent until reaching a circle, with a *block* consisting of eight trials. The subjects were instructed to take small breaks between the blocks when needed. The subject navigated the helicopter in three different navigational *modes*, that alternated in a fixed order. The three navigational modes were then repeated ten times as can be seen in figure 4.3, summing to a total of 240 trials per subject. The order of blocks and modes is illustrated on figure 4.3**D**.

The helicopter simulator was designed specifically for this experiment, and featured three different and independent ways to vary the difficulty. Each difficulty had a hard and an easy setting, giving eight possible combinations of the three types of difficulty. Each block started with a "warm-up" trial and the remaining eight trials consisted of all possible difficulty combinations in random order. The three types of difficulty were: *Angle* difficulty determined the angle between the next circle and the one just

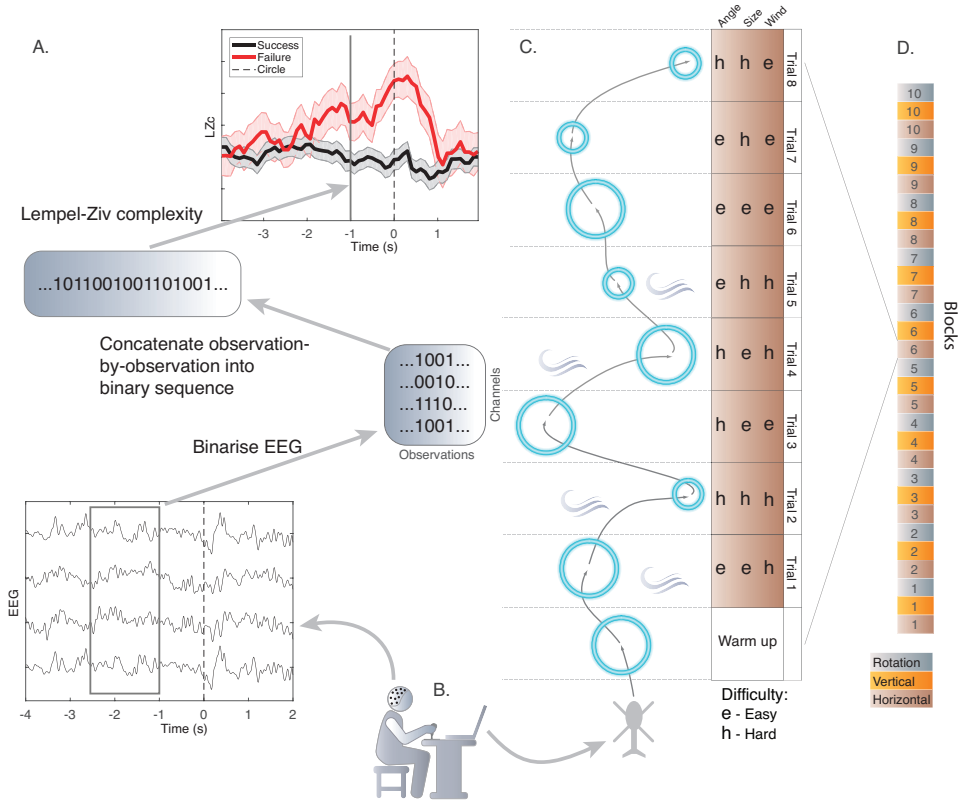


Figure 4.3: Experimental setup and EEG complexity analysis pipeline. **A:** From EEG to LZc. EEG is recorded from a subject interacting with the helicopter simulator (shown in **B**). In windows of 1,500 ms, the EEG is binarised based on the envelope amplitude of each channel. The binary matrix is "stretched" into a string observation-by-observation and given to the Lempel-Ziv algorithm. By using a moving window (with steps of 100 ms) and averaging across trials, we can compare the difference in complexity leading up to a successful or failed navigation. **C:** The objective for the subject was to navigate a helicopter successfully through circles, where a trial is defined as the time leading up to reaching a circle. In total, the subject had to navigate the helicopter through 240 trials divided into blocks of eight trials. Each trial was affected by three independent types of difficulty, each with an easy and a hard setting. Each block of trials contained all eight difficulty combinations in a random order. Trial 4, in the illustrated block, has a hard angle between the circles, the circle is large (easy) and there is turbulence from the hard wind difficulty. **D:** In the helicopter simulator the subject had to control the helicopter in three different navigational modes, where the control scheme differed. The modes were played in a fixed order and repeated ten times. Figure from D.

completed. A hard difficulty therefore forced the subject to take sharper turns with the helicopter. *Size* difficulty determined the size of the circle the helicopter needed to be navigated through, thus a hard difficulty required the subject to have a better and smoother control over the helicopter. *Wind* difficulty decided whether or not there would be turbulence, which made it harder to have a controlled navigation of the helicopter. The difficulties are illustrated on figure 4.3**C**.

4.3.2 EEG acquisition and extra preprocessing

During the experiment, we recorded EEG from subjects using a 64 channel Biosemi system with active electrodes at a sampling frequency of 256 Hz. All subsequent preprocessing and analysis of the EEG was conducted digitally in Matlab.

As we intended to measure the information shared between brain areas, we opted not to employ digital re-referencing to avoid inducing signals between distant electrodes. Therefore, the only interactions between electrodes were with the CMS and DRL electrodes, which are placed in the occipital-parietal area in the Biosemi system.

The subjects used a joystick to interact with the helicopter simulator, which meant that this study contained a higher degree of movement than normally seen in experiments where EEG is recorded. EEG can be sensitive to movement artefacts, which in our experiment could create false positives for difficult trials, where the subject might move their arm more or even unintentionally moving their entire body.

To ensure that movement artefacts didn't influence the LZc, we implemented an aggressive preprocessing using independent component analysis (ICA). Dipole-fitting and automatic classification of the ICs, using plug-ins for the EEGLAB toolbox, was used to help identify non-cortical sources and remove them from the EEG (Delorme and Makeig, 2004; Oostenvelt et al., 2003; Frølich et al., 2015). This preprocessing step was performed between the filtering and the surface Laplacian described in section 4.1.4.

For the analysis we decided to use a subset of eight electrodes from the occipital-parietal area. This subset was selected based on tests from pilot experiments on an early version of the helicopter simulator.

4.3.3 LZc increases during failed trials

With our experiment we wanted to investigate the capacity of EEG complexity to reflect neural changes leading up to failures while navigating a helicopter. To this end, we time-locked the LZc to when the helicopter reached a circle and averaged across trials. On the subject level, we could see a clear difference in the LZc between successful and failed trials, but we also identified a strong difference in subject-specific offsets in the LZc.

For some subjects the lowest measured LZc was still higher than largest LZc from other subjects. Such subject offsets in LZc have been reported in other complexity studies (see e.g. Schartner et al., 2015), and appear unrelated to the experimental conditions. The variation in offsets appeared more pronounced in our study, which

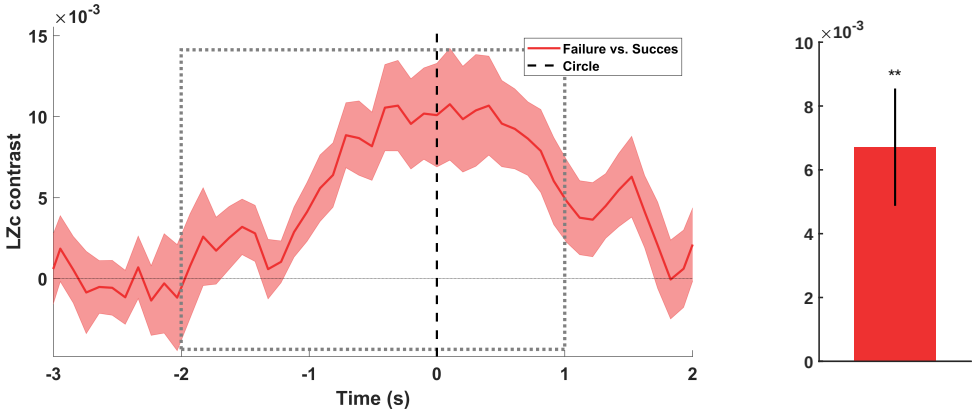
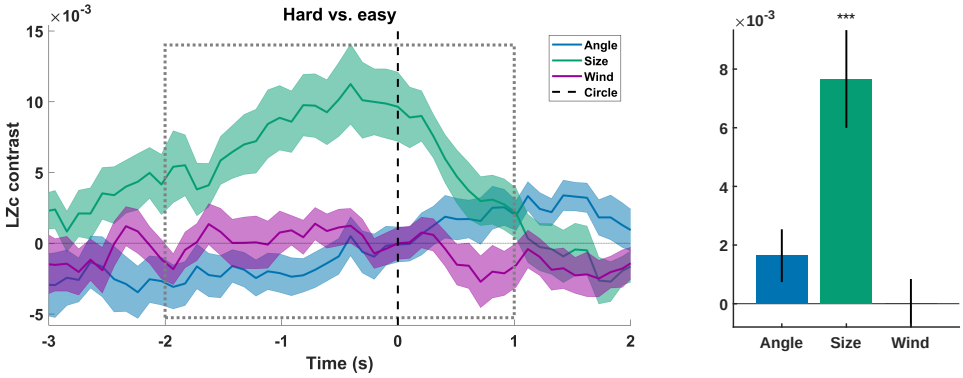


Figure 4.4: (a): LZc time-locked to the time point when the helicopter reached a circle (defined as t = 0 s) for the contrast between failed and successful trials. The LZc contrast was calculated by subtracting the average of successful trials from the average of failed trials for each subject. The contrast was then averaged across subjects, with shaded areas signifying SEM. Note that the two groups of trials were not balanced with respect to the number of trials. The dashed grey square indicates the temporal ROI used to average into single LZc values for each trial. The ROI LZc has been averaged across subjects in **(b)**. Error bars signify the standard error of the subject mean. Figure from D.

might be an effect of the more aggressive preprocessing. We were unsuccessful in removing subject-specific offsets through normalisation, without confounding our experimental paradigm. We solved this by calculating condition contrasts within each subject, which we then averaged across subjects.

For each subject we calculated the average LZc for failed trials and subtracted them by the average LZc from successful trials, resulting in the graph shown in figure 4.4(a). From this contrast it can be seen that the EEG on average is more complex in trials where the subject is about to fail a navigation compared to successful trials. It can also be seen that there is a continued contrast after the circle has been reached. This is likely driven by the 1,500 ms width of the window used to calculate the LZc, which ends in the time point it is designated to. E.g. the LZc designated to $t = 1{,}000$ ms is calculated from the EEG between -500 ms to 1,000 ms relative to reaching a circle.

To test the significance of this contrast we selected a temporal region of interest (ROI) from -2,000 ms to 1,000 ms. For each trial we averaged the LZc in this ROI, excluding LZc samples influenced by an ERP generated immediately after the cirle in the previous trial. We then calculated the subject-averages for the two trial subsets. A paired t-test showed that the LZc in the ROI was significantly higher in failed trails than in successful trials ($p = 0.001$).



(a) (b)

Figure 4.5: (a): LZc time-locked to the time point when the helicopter reached a circle (defined as t = 0 s) for three contrast between difficult and easier trials. The LZc contrasts were calculated by subtracting the average of easy trials from the average of hard trials for each subject. The contrasts were then averaged across subjects, with shaded areas signify SEM. The *Angle*, *Size*, and *Wind* contrasts were calculated using the hard and easy difficulty and the members of the two groups are therefore balanced. The dashed grey square indicates the temporal ROI used to average into single LZc values for each trial, which have been averaged across subjects in (b). Error bars signify the standard error of the subject mean. Figure from D.

4.3.4 LZc is sensitive to increased difficulty

In a real-world scenario, a flight-simulator would in most cases not need neural measures to register, when a navigational failure was made. As in the helicopter-simulator created for this experiment, failures could simply be detected, when the subject failed to complete a task, such as flying through a circle. However, it is much harder for a simulator to register, when the user is struggling with navigating the helicopter without failing. To investigate this, we looked at how the LZc varied with the changes in the three types of difficulty.

In order to examine if changes in LZc was related to the various possible changes in difficulty, we conducted a repeated measures ANOVA with the LZc in the temporal ROI as the response variable and a within-subject design. The ANOVA showed that the subjects, the *Size* difficulty, and the navigational mode had significant effect on the LZc variation ($p_{\text{subject}} < 0.001$, $p_{\text{size}} < 0.001$, $p_{\text{mode}} = 0.0191$), whereas the *Angle* and *Wind* difficulties had low or no effect on the variation in LZc ($p_{\text{angle}} = 0.077$, $p_{\text{wind}} = 0.977$).

Figure 4.5 shows the contrasts between the trials with a hard and an easy setting, calculated in a manner similar to the failure contrast. For example, for the *Angle* difficulty we subtracted the average LZc across all trials with an easy angle from the average of trials with a hard angle. Paired t-tests supports the conclusion of the ANOVA with similar p-values ($p_{\text{angle}} = 0.077$, $p_{\text{size}} < 0.001$, $p_{\text{wind}} = 0.977$).

The temporal waveform of the failure contrast curve on figure 4.4(a) appear similar to the *Size* contrast in figure 4.5(a). This had us asking the question, whether the *Size*

contrast was simply a product of subjects failing the hard trials? To probe this, we divided the trials into two subsets. One subset for all the failed trials, and one for the successful trials. We then calculated the *Size* difficulty contrast for each subset.

We found that for the subset only containing trials with successful navigations, the LZc was still significantly higher when the *Size* difficulty contrast was hard ($p_{\text{size,success}} < 0.001$). This is evidence that LZc not only reflects externally measurable events such as navigational failures, but also reflects more endogenous processes, such as a user successfully completing a task, but struggles with it.

A change in LZc over time, could reflect a learning process taking place, as subjects got more comfortable with the helicopter simulator. We therefore included time, represented by the block number, as a predictive variable in the repeated measures ANOVA. This showed time to have a significant effect on LZc ($p < 0.001$), as well as having a significant interaction with the navigational mode ($p = 0.002$). We therefore investigated the difference of LZc in the temporal ROI between the first and the last block for each navigational mode, looking only at successful trials.

We found a drop in complexity between the first and last block for all three navigational modes. The contrast was strongest for the Horizontal mode, and it was also the only mode to have a significant drop in complexity as measured be a paired t-test ($p = 0.003$). This result corroborated post-experiment questionnaires, where the Horizontal mode was reported by subjects to be the easiest navigational mode. Additionally, the subjects reported that the Horizontal mode was the mode they learned first, using self-scored learning curves.

4.4 Summary of complexity studies

For better or for worse I am an engineer and not a philosopher. I therefore have a rather pragmatic approach to the interpretation of the results presented in this chapter.

4.4.1 EEG complexity as a neural marker

We have shown that LZc correlates with changing stages of sleep, and the ability to process external information during sleep. Furthermore, we found the the EEG complexity prior to a stimulus was significantly correlated to the strength of an ERP marker for motor responses for all sleep stages, but the deep NREM stage. In REM sleep, there was a reversal of correlation between complexity and motor indices, which was interpreted to indicate "drastically different gating mechanisms across sleep stages".

While subjects interacted with our helicopter simulator, LZc was able to distinguish between successful trials and moments leading to failure consistently across subjects. Furthermore, in successful trials LZc was able to contrast the trials based on the difficulty of the task. This suggests that LZc is sensitive to moments where the subject is struggling, perhaps due to a higher mental workload.

Combined the results support studies reporting EEG complexity's capability at indexing states of consciousness. Additionally our results show the promises of using complexity as an implicit marker of performance.

4.4.2 What does EEG complexity reflect?

In Koch et al. (2016), the authors have a model for the relationship between complexity and consciousness. In this model high complexity is related to high degrees of causal cortical integration and/or high spatial differentiation.

However, the model is based on their PCI variant of complexity, where the main source of information is externally and artificially introduced through TMS. Furthermore, the PCI is calculated by averaging over several TMS events, thereby averaging out endogenous processing. Therefore, the PCI mainly looks at how the induced information is integrated throughout the brain.

The ratio between the TMS and other signals (both observation noise and endogenous neural processing) is very high, essentially meaning that there is no activity in areas that does not react to the TMS pulse. So a decrease in integration means that a larger area of cortex shows no activity. This causes the averaged EEG to be more compressible and the complexity to decrease.

This model is therefore not easy to extend to complexity calculated from EEG without TMS stimulation. One reason is that the model does not take into account that in states of high awareness it is likely that multiple sources of activity contribute to the EEG.

I propose that complexity reflects something else for spontaneous EEG. In a state of high neural activity, where multiple sources contribute to the EEG, a high degree of integration could decrease the complexity. With many active sources of activity, an increase in integration across the brain could mean that a single, or few, sources dominated the recorded EEG. This would make the topographical information more compressible and lead to a decrease in complexity.

This view is supported by data from both studies presented in this chapter. By using short time windows with a high overlap, we ensured a high temporal resolution that enabled us to study evoked complexity. Figure 4.2b shows a large drop in LZc immediately after a sound has been played, eliciting an ERP due to information being propagated through the brain.

For the subjects using the helicopter simulator a similar ERP-like drop in LZc could be seen. For some individual subjects a sudden decrease in LZc occurred immediately after navigating through a circle (see figure S1 in D).

This model for spontaneous complexity could collaborate well with the global neuronal workspace (GNW) model (Dehaene and Changeux, 2011). The GNW model describes that when we have a conscious experience of a stimulus, the related information is broadly shared and broadcasted throughout the cortex. Conscious perception would cause a single information stream to dominate the EEG by being integrated through the cortex. This would hypothetically cause a larger decrease in complexity, compared to the lower integration of subliminal perception.

Thereby EEG complexity can be seen as a measure of cognitive integration, with a possible link to theories of consciousness.

APPENDIX A

EEG in the classroom: Synchronised neural recordings during video presentation

Authors: **Andreas Trier Poulsen**, Simon Kamronn, Jacek Dmochowski, Lucas C. Parra, and Lars Kai Hansen.

Status: Published in *Scientific Reports.*

URL: http://www.nature.com/articles/srep43916

Supplementary materials included.

SCIENTIFIC REPORTS

OPEN

EEG in the classroom: Synchronised neural recordings during video presentation

Received: 26 April 2016
Accepted: 01 February 2017
Published: 07 March 2017

Andreas Trier Poulsen[1,*], Simon Kamronn[1,*], Jacek Dmochowski[2,3], Lucas C. Parra[3] & Lars Kai Hansen[1]

We performed simultaneous recordings of electroencephalography (EEG) from multiple students in a classroom, and measured the inter-subject correlation (ISC) of activity evoked by a common video stimulus. The neural reliability, as quantified by ISC, has been linked to engagement and attentional modulation in earlier studies that used high-grade equipment in laboratory settings. Here we reproduce many of the results from these studies using portable low-cost equipment, focusing on the robustness of using ISC for subjects experiencing naturalistic stimuli. The present data shows that stimulus-evoked neural responses, known to be modulated by attention, can be tracked for groups of students with synchronized EEG acquisition. This is a step towards real-time inference of engagement in the classroom.

Engagement and attention are important in situations of learning, but most methods for measuring of attention or engagement are intrusive and unrealistic in everyday situations[1–3]. Recently, inter-subject correlation (ISC) of electroencephalography (EEG) has been proposed as a marker of attentional engagement[4–6] and we ask in this work whether it can be recorded robustly with commercial-grade wireless EEG devices in a classroom setting. Furthermore, we address two other issues related to the robustness of the signal: The potential neurophysiological origin of the measure and the robustness of the detection scheme to inter-subject variability in spatial alignment.

User engagement has been defined as '… the emotional, cognitive and behavioural connection that exists, at any point in time and possibly over time, between a user and a resource'[7]. Traditional approaches to measuring engagement are based on capturing user behaviour via user interfaces, self-report, or manual annotation[8]. However, tools from cognitive neuroscience are increasingly being employed[9]. Recent efforts in neuroscience aim to elucidate perceptual and cognitive processes in a more realistic setting and using naturalistic stimuli[4,10–14]. From an educational perspective such quantitative measures may help identify mechanisms that make learning more efficient[9], align services better with students needs[7], or monitor critical task performance[15]. The potential uses of engagement detection in the classroom are numerous, e.g., real-time and summary feedback for the teacher, motivational strategies for increased student engagement, and screening for impact of teaching materials. Before the findings of tracking attentional responses with neural activity[4–6] can be employed in a real-time classroom scenario, several issues must be addressed first, including: (1) Is it possible to reproduce the ISCs to naturalistic stimuli under the adverse conditions of a classroom? (2) Are the ISCs robust to inter-student variability of the spatial information processing networks? And (3) can ISCs be recorded with equipment that is both comfortable and affordable enough to make it a realistic technology for schools?

Here we investigate the feasibility of recording such neural responses from students who are viewing videos. We use an approach developed by Dmochowski *et al.*[4] that uses inter-subject correlation (ISC) of EEG evoked responses. The basic premise is that subjects who are engaged with the content exhibit reliable neural responses that are correlated across subjects and repetitions within the same subject. In contrast, a lack of engagement manifests in generally unreliable neural responses[6]. ISC of neural activity while watching films have been shown to predict the popularity and viewership of TV-series and commercials[5], and shows clinical promises as a measure of consciousness levels in non-responsive patients[16] (fMRI study). We argue here that the neural reliability of students indeed may be quantified on a second-by-second basis in groups and in a classroom setting, and we seek

[1]Technical University of Denmark, DTU Compute, Kgs. Lyngby, Denmark. [2]Stanford University, Department of Psychology, Palo Alto, USA. [3]City College of New York, Department of Biomedical Engineering, New York, USA. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.T.P. (email: atpo@dtu.dk)

Figure 1. Experimental setup for joint viewings. (**Left**) 9 subjects where placed on a line to induce a cinema-like experiences. (**Right**) Subjects seen from the back, watching films projected onto a screen. Tablets recording EEG are resting on the tables behind the subjects. The signal is transmitted wirelessly from each subject.

to investigate the robustness of measuring it with electroencephalography (EEG) responses during exposure to media stimuli.

To enable correlations between multi-dimensional EEG, correlated component analysis (CorrCA) was introduced[4]. CorrCA finds multiple spatial projections that are shared amongst subjects, such that their components are maximally correlated across time. Here we are interested in the reproducibility of using CorrCA as a measure of inter-subject correlation, and will focus predominantly on the first component, which captures most of the neural responses shared across students.

The main goal of the present work is to determine whether student neural reliability can be quantified in a real-time manner based on recordings of brain activity in a classroom setting using a low-cost, portable EEG system – the Smartphone Brain Scanner[17]. With regard to the robustness of the detection scheme, we report on both theoretical and experimental investigations. First, we show that ISC evoked by rich naturalistic stimuli is robust enough to be reproduced with commercial-grade equipment, and to be recorded simultaneously from multiple subjects in a classroom setting. This opens up for the possibility of real-time estimation of student attentional engagement. Secondly, we show mathematically that the CorrCA algorithm is surprisingly robust to variations in the spatial patterns of brain activity across subjects. Finally, we demonstrate that the level of ISC is related to a very basic visual response that is modulated by narrative coherence of the video stimulus.

Results

To monitor neural reliability we used video stimuli as they provide a balance between realism and reproducibility[11]. We recorded EEG activity using the Smartphone Brain Scanner while subjects watched short video clips of approximately 6 minutes duration, either individually or in a group setting (Fig. 1). To measure reliability of EEG responses, we used correlated components analysis (CorrCA, see Methods) to extract maximally correlated time series with shared spatial projection across repeated views within the same subject (inter-viewing correlation, IVC), or between subjects (inter-subject correlation, ISC).

One of our main points of interest is to investigate the robustness of ISC from EEG recorded in a classroom through comparisons with results previously measured in a laboratory setting[4]. We therefore employed similar methods of analysis and calculated ISCs and IVCs in 5 second windows with 80% overlap to investigate their temporal development in a 1-second resolution. We chose to analyse the EEG with CorrCA in a broad frequency band (0.5 and 45 Hz), instead of investigating specific frequency bands, to keep the analysis methods comparable with the prior lab-based study. Moreover, CorrCA is a method used for robustly measuring ISC with low computational costs; hence making it a good candidate for long term real-time analyses on small devices in a classroom setting.

The subjects watched three video clips, which were presented twice in random order. The first video was a suspenseful excerpt from the short film, *Bang! You're Dead*, directed by Alfred Hitchcock. It was selected because it is known to effectively synchronize brain responses across viewers[4,18]. The second video was an excerpt from *Sophie's Choice*, directed by Alan J. Pakula (1982), and the third was an uneventful baseline video of people silently descending an escalator. For both the joint and individual recording scenarios, the time course of the ISC, based on the first CorrCA component from subjects watching the film, closely reproduces results obtained previously in a laboratory setting (Fig. 2a and Table 1).

An indication of the stability of the technique is provided by the spatial patterns of the neural activity that drives these reproducible responses. Similar to other component extraction techniques, such as independent component analysis or common spatial patterns[19,20], CorrCA reduces the signal of multiple electrodes to a few components. The ISC is then computed for the first few components, which capture most of the correlation between recordings. The strongest three correlated components show a stable pattern of activity across the different groups and recording conditions (Fig. 2b), all three obtaining significant spatial correlations between groups ($r_{comp1} = 0.97$, $r_{comp2} = 0.91$, $r_{comp3} = 0.79$, all with $p < 0.002$ for uncorrected permutation test), for *Bang! You're Dead*. The robustness to recording conditions is also apparent for the second film clip from *Sophie's Choice* ($r_{comp1} = 0.51$, $p < 0.002$; $r_{comp2} = 0.48$, $p = 0.008$; $r_{comp3} = 0.36$, $p = 0.033$), albeit with a lower average correlation, which for the first two components may be due to noisy scalp maps for the *Joint 1* group and *Individual* group,

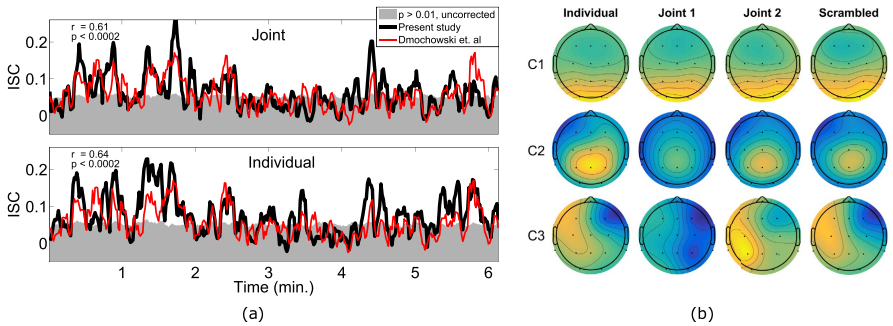


Figure 2. ISC of neural responses to naturalistic stimuli are robust across different groups of subjects and reproducible in a classroom setting. (**a**) Comparison between the ISC obtained by Dmochowski *et al.*[4] and the present study for the first CorrCA component and the first viewing of *Bang! You're Dead*. The ISC is calculated with a 1-second resolution (5 s windows, 80% overlap). The grey area indicates chance levels for ISC ($p > 0.01$ estimated with time-shuffled surrogate data, uncorrected for multiple comparisons). (**b**) The corresponding scalp projections of the first three components obtained from the correlated component analysis (CorrCA) of each of the four subject groups watching *Bang! You're Dead* the first time. For each component, CorrCA finds one shared set of weights for all subjects in the group. Four distinct groups of subjects watched videos in different scenarios: individually on a tablet computer (*Individual*), individually with order of scenes scrambled in time (*Scrambled*), and jointly in a classroom as seen in Fig. 1 (*Joint 1* and *Joint 2*). For each projection, the polarity was normalized so the value at the Cz electrode is positive.

	ISC v1	ISC v2	IVC
Individual	0.64**	0.33**	0.49**
Joint group 1	0.51**	0.15**	0.44**
Joint group 2	0.61**	0.28**	0.54**

Table 1. Correlation coefficients between the ISC time courses obtained in a laboratory setting[4] and those obtained in the present study (groups *Individual, Joint 1* and *Joint 2*). Inter-subject correlation (ISC) measures similarity of responses between subjects for the first and second viewings (v1, v2), and the inter-viewing correlation (IVC) measures similarity within-subject between the two views. Coefficients are calculated for the first CorrCA component recorded while watching *Bang! You're dead*. $^{**}p < 0.01$.

respectively (see Supplementary Fig. S1). For the baseline video, only the first component achieved significant average correlation between groups ($r_{comp1} = 0.46$, $p = 0.014$). The lower stability in the scalp maps obtained for *Sophie's Choice* and the baseline video could be explained by the lower ALD of these stimuli (see below), since these films obtain lower average IVC compared to *Bang! You're Dead* for all groups (Fig. 3).

Previous research has indicated the potentials of ISC as a marker of engagement of conscious processing[4,5,6,12,16]. To further investigate this, we asked subjects post-experiment to describe the film segments (or "scenes") that made the biggest impact on them. We quantified their answers by assigning each answer to one of eight general scene descriptions. Table 2 shows that the scenes most frequently mentioned are "Boy pointing gun at mother" or "Boy pointing gun at people", and 29 out of 30 subjects mentioned one or both of the scenes as having had high impact on them. The most frequently mentioned scene occurs around 2:25, where a peak in the ISC can be seen (Fig. 2a). The high impact of this particular scene was confirmed by the suspense ratings presented in Naci *et al.*[16]. See Dmochowski *et al.*[4] for additional descriptions and examples of scenes eliciting high ISC in *Bang! You're Dead*.

To determine if the portable equipment, which uses only 14 channels, can detect varying levels of neural reliability, a second group of subjects watched the same two film clips individually, but now with scenes scrambled in time. This intervention is a widely used tool to create a baseline with similar low-level stimuli, yet reduced engagement[4,18,21,22]. See Methods for more information on the definition and time scales of the scrambled scenes. Despite using consumer-grade EEG we find that IVC is significantly above chance for a large fraction of the original engaging clip, but drops dramatically when the scenes are scrambled in time (mean IVC, Fig. 3, $p < 0.01$, for *Bang! You're Dead*). Also the baseline video, which subjects reported not to engage them at all, only obtained significant ISC ($p < 0.01$, uncorrected) in 2.3% of the 354 tested time windows, compared to the 54.1% significant windows obtained for *Bang! You're Dead*.

For experiments conducted in less controlled, everyday settings as in this study, it is important to assess across-session reproducibility. To test this, we recorded a second group of subjects in a classroom setting who watched the material together (*Joint 1* and *2*). These two groups obtained mean IVCs comparable to the individual recordings (Fig. 3, *Bang! You're Dead*: $p > 0.49$, *Sophie's Choice*: $p > 0.26$), and also showed reproducibility between the groups of simultaneous recordings (Fig. 3, *Bang! You're Dead*: $p > 0.49$, *Sophie's Choice*: $p > 0.08$).

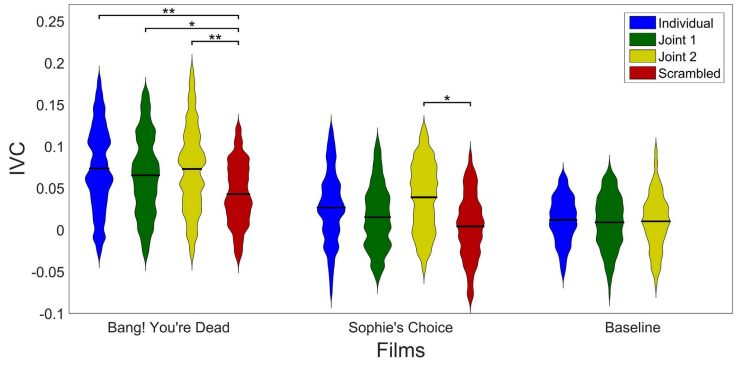


Figure 3. Distribution and mean of IVC calculated from the first CorrCA component for subject groups and films. Violin plots show distributions of IVC estimated using a squared exponential (normal) kernel with bandwidth of 0.005[41]. Horizontal black bars denote distribution means. For visualisation purposes, the extreme 2.5% values at either end of the distributions were left out of the violin plots (but were kept for estimating mean and p-values). A block permutation test (block size $B = 25\,s$) was employed to estimate statistical significant differences in the mean IVC between viewing conditions (uncorrected for multiple comparisons). For both films there were significant differences in mean IVC between groups with normal narrative and the *Scrambled* group (*Bang! You're Dead*: $p_{Individual} = 0.006$, $p_{Joint1} = 0.033$, $p_{Joint2} = 0.004$; *Sophie's Choice*: $p_{Individual} = 0.059$, $p_{Joint1} = 0.37$, $p_{Joint2} = 0.012$). However, there were no significant differences between groups with the original, unscrambled narrative. Note that the *Scrambled* group did not watch the baseline video.

Scene	Approx. times	No of times mentioned (%)
The boy shoots (or points gun at) mother	2:25 and 3:00	16 (53%)
The boy shoots (or points gun at) at people	2:10, 3:30 and 5:30	15 (50%)
The boy loads another bullet into gun	6:10	8 (27%)
The uncle discovers his gun is gone	4:35	4 (13%)
The boy finds and loads gun	0:25 and 1:40	4 (13%)
The boy points at mirror or shoot towards camera	0:40, 1:50 and 5:25	4 (13%)
When the father did not run after the boy	3:00	1 (3%)
The abrupt ending	6:14	1 (3%)

Table 2. Scenes described by the subjects as having the strongest impression on them. Based on the 30 subjects which saw *Bang! You're Dead* with uninterrupted narrative. In a post-experiment questionnaire, subjects were asked to describe the scenes that made the strongest impression on them. Their answers were collected in the eight groups. The subjects each mentioned 1.77 scenes on average (0.77 std.). 29 subjects (97%) mentioned either scenes where the boy points the gun at his mother or at other people.

Robustness to inter-subject variations in the spatial brain structure is a basic question when applying CorrCA to classroom data. CorrCA is derived under the assumption that the spatial networks of subjects are identical. This assumption could be challenged by inter-individual differences, however, it turns out to be surprisingly robust to such variability[23]. To demonstrate this, we briefly analyse a 'worst case' scenario in which the true mixing weights of two subjects form a pair of *orthogonal* vectors. The observations are assumed to consist of a single true signal, $\mathbf{z}$, mixed into D dimensions with additive Gaussian noise; $\mathbf{X}_1 = \mathbf{a}_1\mathbf{z}^\top + \varepsilon$, $\mathbf{X}_2 = \mathbf{a}_2\mathbf{z}^\top + \varepsilon$. Given a large sample, the covariance matrices are given as $\mathbf{R}_{11} = P \cdot \mathbf{a}_1\mathbf{a}_1^\top + \sigma^2\mathbf{I}$, $\mathbf{R}_{12} = P \cdot \mathbf{a}_1\mathbf{a}_2^\top$, where P is the variance of $\mathbf{z}$ and σ^2 signifies the noise variance. For simplicity the weight vectors are assumed to be unit length. The two matrices in Eq. (3) can then be written as

$$(\mathbf{R}_{11} + \mathbf{R}_{22})^{-1} = \frac{1}{P}\left([\mathbf{a}_1\,\mathbf{a}_2]\begin{bmatrix}\mathbf{a}_1^\top \\ \mathbf{a}_2^\top\end{bmatrix} + \frac{2\sigma^2}{P}\mathbf{I}\right)^{-1} \;;\; \mathbf{R}_{12} + \mathbf{R}_{21} = P \cdot [\mathbf{a}_1\,\mathbf{a}_2]\begin{bmatrix}\mathbf{a}_2^\top \\ \mathbf{a}_1^\top\end{bmatrix}, \tag{1}$$

using block matrix notation. With $\mathbf{a}_1^\top\mathbf{a}_2 = 0$, $\|\mathbf{a}_1\|^2 = \|\mathbf{a}_2\|^2 = 1$ and the Woodbury identity, the product of the two matrices in Eq. (1) can be expressed as

$$(\mathbf{R}_{11} + \mathbf{R}_{22})^{-1}(\mathbf{R}_{12} + \mathbf{R}_{21}) = \frac{P}{2\sigma^2 + P}(\mathbf{a}_1\mathbf{a}_2^\top + \mathbf{a}_2\mathbf{a}_1^\top). \tag{2}$$

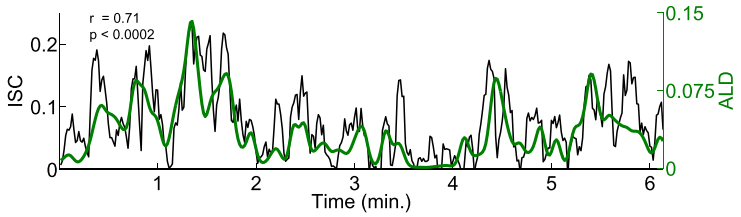


Figure 4. The ISC of the first CorrCA component is temporally correlated with the average luminance differences (ALD) of the film stimulus. ALD is calculated as the frame-to-frame difference in pixel intensity, smoothed to match the 5 s window of ISC, and mainly reflects the frequency of changes in camera position. Data computed from the neural responses of subjects watching *Bang You're Dead*.

	ISC v1	ISC v2	IVC
Bang You're Dead	0.71**	0.61**	0.56**
Sophie's Choice	0.50**	0.24**	0.23**
Bang You're Dead (Scr)	0.54**	0.45**	0.35**
Sophie's Choice (Scr)	0.42**	0.01	−0.22**

Table 3. Correlation coefficients between the ALD and the ISC for the two viewings (v1, v2) as well as the IVC for the first correlated component. The correlation is presented for *Bang You're dead* and *Sophie's Choice* for the *Individual* and *Scrambled (Scr)* groups. $^{**}p < 0.01$.

An eigenvector of matrix (2) takes the form $\alpha\mathbf{a}_1 + \beta\mathbf{a}_2$, with $\alpha = \pm\beta$ and $\pm\frac{P}{2\sigma^2 + P}$ as eigenvalues. By applying this eigenvector to observations, $\mathbf{X}_1$ and $\mathbf{X}_2$, we see that CorrCA still identifies the relevant time series, $\mathbf{z}$.

For the first CorrCA component, the channels weighted most heavily are the ones positioned over the occipital lobe (see Fig. 2b). To estimate how much of the ISC was driven by basic low-level visual processing, we analysed the relation between ISC and a measure of frame-to-frame luminance fluctuations (average luminance difference, ALD; see methods). Note that to avoid synchronised eye artefacts and to ensure that only signals of neural origin contributed to the measured correlations, we removed independent components related to eye artefacts from the EEG (see methods).

Figure 4 and Table 3 show that there is a significant correlation between the ISC and the ALD for both *Bang! You're Dead* and *Sophie's Choice* for the first CorrCA component. This suggests that this portion of the correlated activity may indeed be driven by low-level visual evoked responses. However, the degree of engagement, here represented by narrative coherence, appears to modulate the *amplitude* of the ISC time course, since even though the scrambled stimulus was driven by the visual stimulus, it was so to a lesser extent. Previous research has shown that visual evoked potentials (VEP) are modulated by spatial attention[24] and that even feature-specific attention enhances steady-state VEPs[25]. We quantify the effect of scrambling the narrative by comparing the sensitivity (slope) of ISC to ALD in both the normal and scrambled conditions by fitting a simple linear model (Fig. 5). For both films we found significant reductions of the ISC/ALD slope in the scrambled version ($p < 0.01$; block permutation test, with block size $B = 25\,s$).

Discussion

We have demonstrated that student neural reliability to media stimuli may be quantified using EEG in a classroom setting. For educational technology cost and robustness are key features, hence, we aimed at establishing a realistic scenario based on low-cost consumer grade equipment, the Smartphone Brain Scanner, focusing on several potential sources that could degrade robustness.

We have provided evidence that salient aspects of the neural reliability previously detected with laboratory grade equipment can be reproduced in a realistic setting. We recorded fully-synchronized EEG with nine subjects in a real classroom and found that the level of neural response reliability matched prior laboratory results. The robustness of CorrCA and ISC is granted by the reproducibility between recording conditions, both of the ISC time-courses throughout the film clips and of the spatial topographies of the first three CorrCA components. For the film clip from *Bang! You're Dead* we saw that seven subjects were enough to obtain stable topographies for all three components, whereas for *Sophie's Choice* and the baseline video the results were more noisy, suggesting that more subjects are needed to obtain stable results. Previous research shows that ten subjects provided for stable results in a case involving non-narrative baseline videos or films with lower ISC and IVC in a laboratory setting[4].

Mathematically, we have shown that our detection scheme, CorrCA, is robust to inter-subject variability in spatial configurations of brain networks, or induced by cap misalignment. In the calculations, we assumed two subjects in a worst case scenario where the subjects' spatial projections are orthogonal. This result conforms well with simulations that show that, even for multiple subjects with randomly drawn spatial projections, CorrCA was able to find the relevant times series[23]. The simulations also showed that increasing the number of subjects

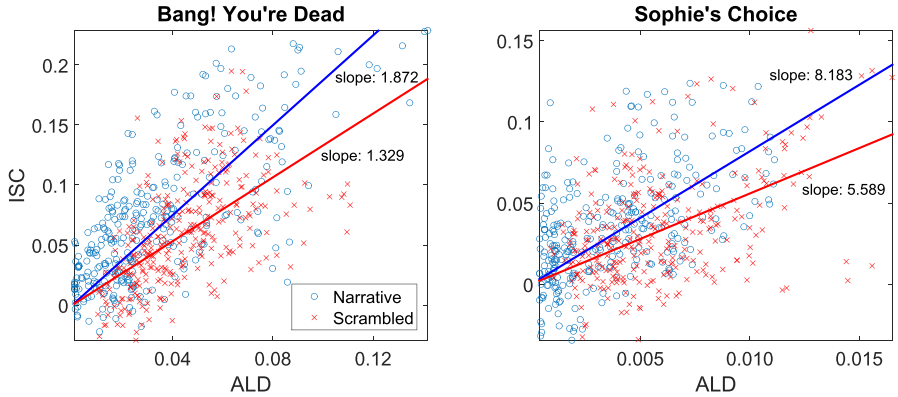


Figure 5. Relation between the ISC and the ALD for different conditions. Each point indicates a point in the ISC time course as seen in Fig. 2a (5 s windows, 80% overlap) and the corresponding ALD calculated from the visual stimulus. It is evident that time points with higher luminance fluctuations (hight ALD) result in higher correlation of brain activity across subjects (high ISC). The indicated "slope" is a least squares fit of the slope of lines passing through (0, 0). The slope indicates the strength of ISC for a given ALD value. For both films there is a significant drop in the slope ($p < 0.01$: block permutation test with block size $B = 25$ sec), thus the original narrative (blue) elicits higher ISC than the less engaging scrambled version of the films (red). Note that brightness of the scenes in *Sophie's Choice* is much lower than in *Bang! You're dead*, resulting in an ALD that is lower by almost a factor 10.

decreased the signal-to-noise ratio, presumably due to the estimated common projection not being able to fit with the different projections of each subject.

We have presented results that further indicate a relationship between changes in ISC and viewer engagement. Through a basic analysis of questionnaires on scenes of high impact, we found that high ISC indeed is associated with high impact. We have also showed a relationship between neural responses to luminance fluctuations and coherence of stimulus narrative. For both the films presented, we saw a significant drop in the average IVC for subjects watching the film sequences in which the narrative had been temporally scrambled. At the same time no significant difference was found between the groups watching the film sequences that had not been scrambled, which further underlines the robustness of the measure.

It may appear surprising that there exists a significant correlation between the *raw EEG signals* of various students in the classroom. However, it is well-known that eye scan patterns in a film audience follow a specific pattern after a scene change, activating the dorsal pathway[26]. A valid assumption could therefore be that the correlation is due to synchronised artefacts from eye movements, but this has recently been shown not to affect attentional modulation of ISC[6]. Also, it is known that stimuli in the form of flashing images elicit VEPs, which are modulated in amplitude by the luminance[27]. When recorded with EEG, the spatial distribution of the early VEP at 100 ms (P100) is similar to the scalp maps of the first correlated component (C1 in Fig. 2b)[24,28].

We investigated whether low-level visual processes could be a driving force behind the measured ISCs by correlating the ISC with changes in luminance in the video stimuli, as measured by the ALD. We found that luminance fluctuations drive a significant portion of the ISC.

In all four groups of subjects *Sophie's Choice* obtained lower IVC compared to *Bang! You're Dead*. This difference could be explained by the fact that the film clip also had a much lower ALD. Also, Fig. 4 indicates that the passage in *Bang! You're dead* with the highest and most sustained ISC (around 1:20 to 1:50) coincides with the interval with the most scene changes. This relationship could, however, also be due to more complex processes, as fast-paced cutting is a known cinematographic tool used by Hitchcock to induce suspense and thereby increase the attention of the viewer[29].

The strong link between ISC and luminance fluctuations due to scene cuts has also recently been presented in a fMRI study[30]. This is something that would be interesting to take into account for future studies investigating the applicability of ISC. Baseline videos could be created in ways to achieve similar ALD features as the target stimuli. The baseline video, created for this study, consisted of one continuous scene of people entering and exiting an escalator in a relaxed manner, which did not produce any significant correlation. Future studies might use a baseline video containing scene cuts of faces and body parts, to also take the effect of editing into account.

To investigate the possibility of higher level processes also being at play, we analysed the linear relationship between ISC and luminance fluctuations at a given time in the video stimulus. The scrambling operation aimed to test for a change in attentional engagement while controlling for low level features. The premise was that subjects would be less attentive to the stimulus, i.e. less "engaged", if they did not follow the narrative arch of the story. With that in mind, Figs 4 and 5 suggest that ISC is driven by stimulus-evoked responses that are modulated by attentional engagement with the stimulus.

We have demonstrated the feasibility of tracking inter-subject correlation in a classroom setting; a measure that has been related to attentional modulation[6]. We have shown that ISC is robust to recording equipment and

conditions, and we have presented evidence that the amplification of ISC in films that have a strong and coherent narrative is due to attentional modulation of visual evoked responses. Thus ISC may be used as an indirect electrophysiological measure of engagement through an attentional top-down modulation of low-level neural processes. Recent research has shown that attentional modulation of neural responses takes place in speech perception[31,32], which lends credibility to a similar process occurring in the visual system. The evidence that such a basic and well defined mechanism could be at play further adds to the robustness of the approach in real everyday scenarios.

Methods

Protocol. Four groups of subjects watched the video stimuli in different scenarios. The first group ($N = 12$, *Individual*) watched videos individually in an office environment on a tablet computer (Google Nexus 7 tablet, with a 7″ (17.8 cm) screen) with earphones. The second group ($N = 12$) saw the videos in the same manner, but the scenes of the film stimulus were scrambled in time resulting in the narrative being lost (*Scrambled*). The objective of this condition was to demonstrate that the similarity of responses across subjects is not simply the result of low-level stimulus features (which are identical in the *Individual* and *Scrambled* conditions), but instead, is modulated narrative coherence, which presumably engages viewers. Two additional groups ($N = 9$, $N = 9$) watched the original videos on a screen in a classroom (Fig. 1, *Joint 1* and *Joint 2*), with sound projected through loudspeakers. An attempt was made to create viewing conditions for the subjects in the *joint* groups, that were similar to the viewing conditions for the *individual* group, i.e., lights were dampened and the projected image produced approximately the same field-of-view (see Supplementary materials). The central question was whether the viewing condition (i.e., in a group versus individually) influences the level of ISC across subjects.

Stimuli. The first video clip was a suspenseful excerpt from the short film *Bang! You're Dead* (1961) directed by Alfred Hitchcock. It was selected because it is known to elicit highly reliable brain activity across subjects in fMRI[11] as well as EEG[4]. Our second stimulus was a clip from *Sophie's Choice*, directed by Alan J. Pakula (1982), which has been used earlier to study fMRI activity in the context of emotionally salient naturalistic stimuli[33]. A third non-narrative control video was recorded in a Danish metro station of several people who were being transported quietly on an escalator. Each video clip had a length of approximately six minutes and was shown twice to each subject. For each viewing the order of the clips was randomized, while the same random order was used the second time the clips were shown. A combined video was created for each of the six possible permutations of the order of the clips, starting with a 10 second 43 Hz tone for use in post processing synchronization, and 20 seconds black screen between each film clip. The total length of the video amounted to 39 minutes. An additional control stimulus (*Scrambled*) was created by scrambling the order of the scenes in *Bang! You're Dead* and *Sophie's Choice* in accordance with previous research[4,18]. In these studies, scene segments were defined in varying temporal scales (36 s, 12 s, and 4 s) that consisted of multiple camera positions, "shots". For this study we defined a scene as a single shot (i.e. the segment between two scene cuts) with the added rule that a scene must not exceed 250 frames (~10 s) to reduce subjects' ability to infer the narrative from long scenes. This procedure resulted in 73 scenes lasting between 0.5 and 10 seconds and corresponded to the intermediate to short time-scales employed in previous studies[18].

Subjects. A total of 42 female subjects (mean age: 22.4 y, age range: 18–32 y), who gave written informed consent prior to the experiment, were recruited for this study. Non-invasive experiments on healthy subjects are exempt from ethical committee processing by Danish law[34]. Among the 42 recordings, nine were excluded due to unstable wireless communication that precluded proper synchronization of the data across subjects (five from the *Individual* group, one from the *Scrambled* group and three from the two *Joint* groups). The difference in the number of recordings in the different groups could give unfair advantages with respect to noise when using CorrCA or calculating ISC. We therefore decided to randomly choose four subjects from the *Scrambled* group and one from *Joint 2* group and excluded these from the analyses. This was to ensure that each group had seven fully synchronized recordings.

Portable EEG – Smartphone Brain Scanner. Research grade EEG equipment is costly, time-consuming to set up, and immobile. However, recently consumer grade EEG equipment that is more affordable and has increased comfort has appeared. Here we use the modified 14 channel system, 'Emocap', based on the EEG Emotiv EPOC headset. For details and validation, see refs 17 and 35. In this study it was implemented on Asus Nexus 7 tablets. An electrical trigger and associated sound was used to synchronize EEG and video signals in the individual viewing condition, while a split audio signal (simultaneously feeding into microphone and EEG amplifiers) was used to synchronize the nine subjects EEG recordings and the video in the joint viewing condition (see Supplementary materials for further information on synchronisation). The resulting timing uncertainty was measured to be less than 16 ms. The EEG was recorded at 128 Hz and subsequently bandpass filtered digitally using a linear phase windowed sinc FIR filter between 0.5 and 45 Hz and shifted to adjust for group delay. Eye artefacts were reduced with a conservative pre-processing procedure using independent component analysis (ICA), removing up to 3 of the 14 available components (Corrmap plug-in for EEGLAB[36,37]).

Correlated component analysis to measure ISC and IVC. CorrCA was presented in Dmochowski *et al.*[4], as a constrained version of Canonical Correlation Analysis (CCA). CorrCA seeks to find sets of weights that maximises the correlation between the neural activity of subjects experiencing the same stimuli. For each neural component, CorrCA finds one shared set of weights for all subjects in the group.

Given two multivariate spatio-temporal time series (termed "view" in CorrCA), $\{\mathbf{X}_1, \mathbf{X}_2\} \in \mathbb{R}^{D \times N}$, with D being the number of measured features (EEG channels) in the two views and N the number of time samples,

CCA estimates weights, $\{\mathbf{w}_1, \mathbf{w}_2\}$, which maximize the correlation between the components, $\mathbf{y}_1 = \mathbf{X}_1^\top \mathbf{w}_1$ and $\mathbf{y}_2 = \mathbf{X}_2^\top \mathbf{w}_2$. The weights are calculated using two eigenvalue equations, with the constraint that the components belonging to each multivariate time series are uncorrelated[38]. CorrCA is relevant for the case where the views are homogeneous, e.g., using the same EEG channel positions, and imposes the additional constraint of shared weights $\mathbf{w} = \mathbf{w}_1 = \mathbf{w}_2$. This assumption can potentially increase sensitivity involving fewer parameters. In CorrCA the weights are thus estimated through a single eigenvalue problem;

$$(\mathbf{R}_{11} + \mathbf{R}_{22})^{-1}(\mathbf{R}_{12} + \mathbf{R}_{21})\mathbf{w} = \rho\mathbf{w}, \tag{3}$$

where, $\mathbf{R}_{ij} = \frac{1}{N}\mathbf{X}_i\mathbf{X}_j^\top$, is the sample covariance matrix[4]. To illustrate the spatial distribution of the underlying physiological activity of the components, we use the estimated forward models ("patterns") as discussed in refs 39 and 40.

Average luminance difference (ALD). Video clips were converted to grey scale (0–255) by averaging over the three colour channels. We then calculated the squared difference in pixel intensity from one frame to the next and took the average across pixels. These signals were non-linearly re-sampled at 1 Hz by selecting the maximum ALD for each 1 *s* interval to emphasise the large differences during changes in camera position (see Figure S2 in Supplementary materials for an comparison between frame-to-frame and smoothed difference). These values were then smoothed in time by convolving with a Gaussian kernel with a "variance" parameter of $2.5\,s^2$. This down sampling and smoothing was aimed at matching the temporal resolution of the ALD to that of the time-resolved ISC computation (5 *s* sliding window with 1 *s* intervals).

Statistical testing. In order to evaluate the statistical relevance of the correlations, we employed a simple permutation test ($P = 5000$ permutations)[4]. To test the robustness of the obtained weights for the spatial projections, we calculated the average correlation of all possible pairings of the four conditions groups for a given component. Again, we employed a permutation test ($P = 5000$ permutations) to evaluate statistical relevance by randomly permuting the channel order for each group and recalculating the average correlation. When testing differences in average IVC between conditions, we used a block permutation test (block size $B = 25\,s$, $P = 5000$ permutations) to account for temporal dependencies.

References

1. Robinson, P. Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning* **47,** 45–99 (1997).
2. Cohen, A., Ivry, R. B. & Keele, S. W. Attention and structure in sequence learning. *J. Exp. Psychol.* [*Learn. Mem. Cogn.*] **16,** 17–30 (1990).
3. Radwan, A. A. The effectiveness of explicit attention to form in language learning. *System* **33,** 69–87 (2005).
4. Dmochowski, J. P., Sajda, P., Dias, J. & Parra, L. C. Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? *Frontiers in human neuroscience* **6,** 112 (2012).
5. Dmochowski, J. P. *et al.* Audience preferences are predicted by temporal reliability of neural processing. *Nature Communications* **5,** 1–9 (2014).
6. Ki, J. J., Kelly, S. P. & Parra, L. C. Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *The Journal of Neuroscience* **36,** 3092–3101 (2016).
7. Attfield, S., Kazai, G., Lalmas, M. & Piwowarski, B. Towards a science of user engagement. In *WSDM Workshop on User Modelling for Web Applications* (ACM International Conference on Web Search And Data Mining, 2011).
8. O'Brien, H. L. & Toms, E. G. Examining the generalizability of the user engagement scale (ues) in exploratory search. *Information Processing & Management* **49,** 1092–1107 (2013).
9. Szafir, D. & Mutlu, B. Artful: adaptive review technology for flipped learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1001–1010 (ACM, 2013).
10. Ringach, D. L., Hawken, M. J. & Shapley, R. Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *Journal of vision* **2,** 2–2 (2002).
11. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject synchronization of cortical activity during natural vision. *science* **303,** 1634–1640 (2004).
12. Lahnakoski, J. M. *et al.* Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage* **100,** 316–24 (2014).
13. Lankinen, K., Saari, J., Hari, R. & Koskinen, M. Intersubject consistency of cortical MEG signals during movie viewing. *NeuroImage* **92,** 217–224 (2014).
14. Chang, W.-T. *et al.* Combined MEG and EEG show reliable patterns of electromagnetic brain activity during natural viewing. *NeuroImage* **114,** 49–56 (2015).
15. Lin, C.-T., Huang, K.-C., Chuang, C.-H., Ko, L.-W. & Jung, T.-P. Can arousing feedback rectify lapses in driving? Prediction from eeg power spectra. *Journal of neural engineering* **10,** 056024 (2013).
16. Naci, L., Sinai, L. & Owen, A. M. Detecting and interpreting conscious experiences in behaviorally non-responsive patients. *NeuroImage* (2015).
17. Stopczynski, A., Stahlhut, C., Larsen, J. E., Petersen, M. K. & Hansen, L. K. The Smartphone Brain Scanner: A Portable Real-Time Neuroimaging System. *PloS one* **9,** e86733 (2014).
18. Hasson, U. *et al.* Neurocinematics: The neuroscience of film. *Projections* **2,** 1–26 (2008).
19. Parra, L. & Sajda, P. Blind source separation via generalized eigenvalue decomposition. *The Journal of Machine Learning Research* **4,** 1261–1269 (2003).
20. Koles, Z. J., Lazar, M. S. & Zhou, S. Z. Spatial patterns underlying population differences in the background eeg. *Brain topography* **2,** 275–284 (1990).
21. Miller, G. A. & Selfridge, J. A. Verbal context and the recall of meaningful material. *The American journal of psychology* **63,** 176–185 (1950).
22. Anderson, D. R., Fite, K. V., Petrovich, N. & Hirsch, J. Cortical activation while watching video montage: An fmri study. *Media Psychology* **8,** 7–24 (2006).
23. Kamronn, S., Poulsen, A. T. & Hansen, L. K. Multiview Bayesian Correlated Component Analysis. *Neural Computation* **27,** 2207–2230 (2015).

24. Johannes, S., Münte, T. F., Heinze, H. J. & Mangun, G. R. Luminance and spatial attention effects on early visual processing. *Cognitive Brain Research* **2,** 189–205 (1995).
25. Müller, M. M. *et al.* Feature-selective attention enhances color signals in early visual areas of the human brain. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 14250–4 (2006).
26. Unema, P. J., Pannasch, S., Joos, M. & Velichkovsky, B. M. Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition* **12,** 473–494 (2005).
27. Armington, J. C. The electroretinogram, the visual evoked potential, and the area-luminance relation. *Vision research* **8,** 263–276 (1968).
28. Sandmann, P. *et al.* Visual activation of auditory cortex reflects maladaptive plasticity in cochlear implant users. *Brain* **135,** 555–568 (2012).
29. Bordwell, D. Intensified continuity: visual style in contemporary American film. *Film Quarterly* **55,** 16–28 (2002).
30. Herbec, A., Kauppi, J. P., Jola, C., Tohka, J. & Pollick, F. E. Differences in fMRI intersubject correlation while viewing unedited and edited videos of dance performance. *Cortex* **71,** 341–348 (2015).
31. Mesgarani, N. & Chang, E. F. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* **485,** 233–236 (2012).
32. Mirkovic, B., Debener, S., Jaeger, M. & De Vos, M. Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. *Journal of Neural Engineering* **12,** 046007 (2015).
33. Raz, G. *et al.* Portraying emotions at their unfolding: a multilayered approach for probing dynamics of neural networks. *NeuroImage* **60,** 1448–61 (2012).
34. Den Nationale Videnskabsetiske Komité. Vejledning om anmeldelse, indberetning mv. (sundhedsvidenskablige forskningsprojekter) (2014).
35. Stopczynski, A. *et al.* Smartphones as pocketable labs: visions for mobile brain imaging and neurofeedback. *International journal of psychophysiology* **91,** 54–66 (2014).
36. Delorme, A. & Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* **134,** 9–21 (2004).
37. Viola, F. C. *et al.* Semi-automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology* **120,** 868–877 (2009).
38. Hardoon, D. R., Szedmak, S. & Shawe-taylor, J. Canonical correlation analysis; An overview with application to learning methods. *Neural computation* **16,** 2639–2664 (2004).
39. Parra, L. C., Spence, C. D., Gerson, A. D. & Sajda, P. Recipes for the linear analysis of EEG. *NeuroImage* **28,** 326–41 (2005).
40. Haufe, S. *et al.* On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage* **87,** 96–110 (2014).
41. Hoffmann, H. violin. m-Simple violin plot using matlab default kernel density estimation (2015).

Acknowledgements

We thank the reviewers for their many constructive comments that greatly improved the manuscript. We thank Ivana Konvalinka, Arek Stopczynski and the DTU Smartphone Brain Scanner team for their assistance and helpful discussions. This work was supported by the Lundbeck Foundation through the *Center for Integrated Molecular Brain Imaging* and by Innovation Foundation Denmark through *Neurotechnology for 24/7 brain state monitoring*.

Author Contributions

A.T.P., S.K., J.D., L.P. and L.K.H. designed research; S.K., A.T.P. and L.K.H. performed research; A.T.P., S.K., J.D., L.P., and L.K.H. contributed analytical tools; A.T.P., S.K., L.K.H. analysed data; A.T.P., S.K., J.D., L.P., and L.K.H. wrote the paper.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Poulsen, A. T. *et al.* EEG in the classroom: Synchronised neural recordings during video presentation. *Sci. Rep.* **7**, 43916; doi: 10.1038/srep43916 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

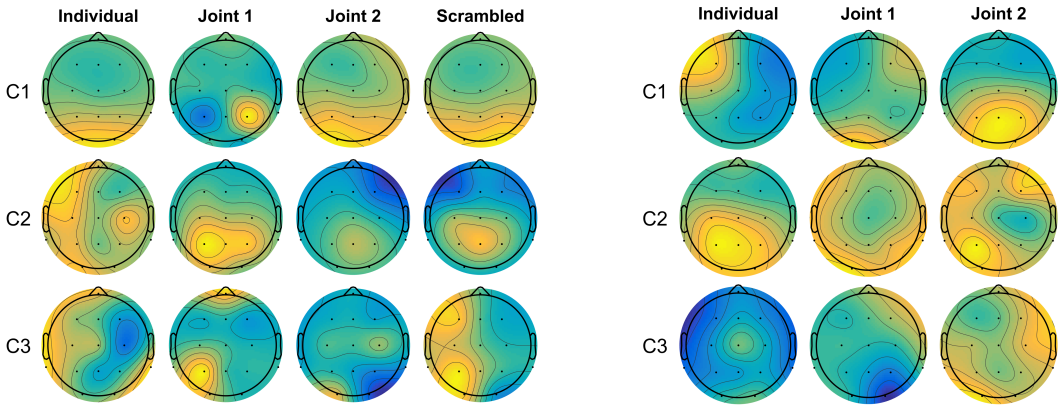
Supplementary Material - EEG in the classroom: Synchronised neural recordings during video presentation

Andreas Trier Poulsen[1,+,*], **Simon Kamronn**[1,+], **Jacek Dmochowski**[2,3], **Lucas C. Parra**[3], **and Lars Kai Hansen**[1]

[1]Technical University of Denmark, DTU Compute, Kgs. Lyngby, Denmark
[2]Stanford University, Department of Psychology, Palo Alto, USA
[3]City College of New York, Department of Biomedical Engineering, New York, USA
[*]atpo@dtu.dk
[+]these authors contributed equally to this work

Contents

S1 Supplementary Tables and Figures



(a) *Sophie's Choice* **(b)** *Baseline video*

Figure S1. The corresponding scalp projections of the first three components obtained from the correlated component analysis (CorrCA) of each of the subject groups watching *Sophie's Choice* and a baseline video for the first time. For each component, CorrCA finds one shared set of weights for all subjects in the group. Four distinct groups of subjects watched videos in different scenarios: individually on a tablet computer (*Individual*), individually with order of scenes scrambled in time (*Scrambled*), jointly in a classroom as seen in Fig. 1 (*Joint 1* and *Joint 2*). For each projection, the polarity was normalized so the value at Cz is positive. Note that the *scrambled* group did not watch the Baseline video.

Table S1. Correlation coefficients between the results obtained in a laboratory setting[1] and those obtained in the present study (groups *Individual, Joint 1* and *Joint 2*). Inter-subject correlation (ISC) measures similarity of responses between subjects for first and second viewings (v1,v2) and the inter-viewing correlation (IVC) measures similarity within-subject between the two views. Values are calculated using the second correlated component recorded while watching *Bang! You're dead*. **: $p < 0.01$, *: $p < 0.05$.

	ISC v1	ISC v2	IVC
Individual	0.05	0.08	0.10*
Joint group 1	0.20**	0.12*	0.28**
Joint group 2	0.27**	0.00	0.10*

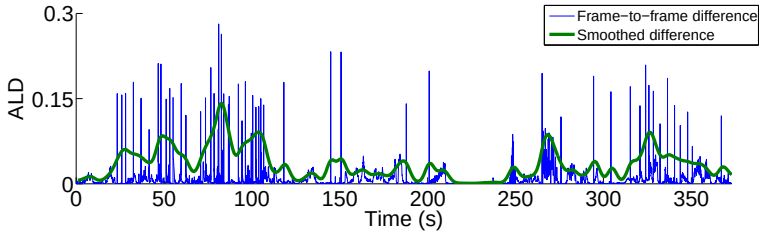


Figure S2. Comparison between the frame-to-frame average luminance difference calculated from *Bang! You're Dead* and the one which have been smoothed with kernel smoothing. Each large value corresponds to a change in camera position.

Table S2. Scenes described as having the strongest impression by subjects. Based on the 30 subjects which saw *Sophie's Choice* with unscrambled narrative. Subjects were asked, in a post-experiment questionnaire, to describe the scenes that made the strongest impression on them. Their answers have here been collected in the nine groups. The subjects mentioned 1.67 scenes on average (0.76 std.). *: Note: only the girl is taken away from her mother, however 47 % of the subjects indicated that both children were taken.

Scene	Number of times mentioned (%)
When both children are taken*	14 (47 %)
The mother is forced to choose between her children	11 (37 %)
The girl is taken from her mother	8 (27 %)
When the girl is screaming at the end	8 (27 %)
When the mother cries at the end	3 (10 %)
Panorama of Jews in lines	2 (7 %)
When the Nazi officer speaks to the mother	2 (7 %)
When I recognised Meryl Streep	1 (3 %)

S2 Experimental setup

To avoid gender having a factor in the results 42 female subjects were recruited with an average age of 22.4 years, distributed with minimum, median, and maximum ages of 18, 22, and 32 respectively. All subjects signed a consent for the use of data, video and image.

The subjects were divided into two groups, with one group of 24 subjects watching the films alone (*Individual viewing*) and another group of 18 subjects subdivided into two groups of nine, which watched the films together (*Joint viewing*). There were taken precautions to ensure that the subjects participating in the same joint viewing, did not know each other beforehand to avoid unwanted confounding factors. The group with individual viewings were additionally evenly divided into a group watching the films with the order of the scenes scrambled and a group watching the film clips normally.

S2.1 Stimulus

One of the goals of the experiments presented here was to recreate the results in ([1]), where the subjects were shown clips from three different films; *Bang! You're Dead* (1961) directed by Alfred Hitchcock, *The Good, the Bad, and the Ugly* (1966), a western directed by Sergio Leone, and a control film of a natural outdoor scene on a college campus. The Hitchcock film produced great results and the same clip was therefore included in the experiments presented here. The western, however, did not produce as many significant times of correlation, and it was decided to replace this clip with one from *Sophie's Choice* (1982) directed by Alan J. Pakula.

The clip from *Sophie's Choice* depicts a young Polish mother on her way to concentration camp during World War II, with her two children. She is accosted by a German officer, who forces her to choose which of her children lives or dies. The dialogue in the film clip is in German. The same film clip was used by ([2]), where the subjects were investigated for emotion-related changes using fMRI and viewer feedback rating. The study found a monotonic increasing response with the

Figure S3. Experimental setup for single viewings.

highest scoring emotions being ”horror”, ”hate”, ”fear”, and ”anger”.

To act as a control, a video was recorded of the escalators Kgs. Nytorv metro station in Copenhagen. This setting was chosen to eliminate the argument that the joint engagement is found for vision of a body versus non-body stimulus. The metro station was chosen as it was rationalised that the passengers getting on the metro in this station, were in less of a hurry compared to other stations, thereby reducing any excitement of people running to catch their train.

Each clip had a length of approximately 6 minutes and were shown twice to each subject. For each viewing the order was randomised, but the same order was used the second time the clips were shown. A combined video was created for each of the six possible permutations of the order of the clips, starting with a 10 second 43 Hz tone for use in post processing synchronisation, and 20 seconds black screen between each film clip. At the end of the video the subject was presented with a text announcing that the video was over, to avoid the subject wondering if they just saw the last clip, between each clip. The total length of the video amounted to 39 minutes.

In ([1]) the order of the scenes in *Bang! You're Dead* were scrambled to investigate the response when the meaning of the film was lost. The same approach was used in this thesis for both *Bang! You're Dead* and *Sophie's Choice*. Since the control video was intended not to carry any meaning, this was left out of the video with scrambled scenes resulting in only two permutations and a length of 23 minutes.

S2.2 Individual viewings

24 subjects were used for the individual viewings, which were conducted in a small office as seen on figure S3. The film was shown on a Google Nexus 7 tablet, with a 7” (17.8 cm) screen with the subject hearing the films through in-ear headphones to avoid wires crossing the head. The headphones had a noise dampening effect which was important due to some of the recordings being made in office hours. Subjects was instructed to sit straight, and avoid movements which could cause artefacts in the EEG, such as chewing, heavy breathing, and limb movement. They were also instructed to look within the frame of the screen to reduce eye artefacts, but was also told to relax and watch the film.

Before the viewing started, each subject drew without replacement for whether the films should be scrambled or not, and afterwards used a dice to decide the order of the film clips.

The subjects were filmed with a camera receiving sound input directly from the tablet as well as from an external microphone. An electric spark was used for post processing synchronisation between the spark showing in the EEG and its clicking sound on the camera recording. As the camera also recorded the sound output from the tablets, the time interval between the spark and the time of the 43 Hz tune could be calculated, and from this the time of start for each film clip.

The lighting in the room was controlled by blacking out the office window and only having an architect lamp on, to ensure the subjects were visible on the camera in the dim light.

S2.3 Joint viewing

The joint viewing experiment is an expansion of the single viewing experiment presented by ([1]). Since recording EEG from nine subjects simultaneously is relatively new territory and presents new obstacles, the experimental setup deviates from the one in the single viewing in some areas.

Different approaches to the placement of the subjects in the room in relation to the screen and to each other were considered. It was decided to go for a ”cinema experience”, with all nine subjects sitting on a line of chairs. By instructing the subjects to keep their eyes within the screen, as in the single viewings, they were not able to directly see the facial expressions of one another. As the films were watched on a projector it was possible to both regulate the distance from the subject to the screen and the length of the diagonal of the picture projected on the screen. It was decided to keep the viewing angle from one corner of the screen to the opposite corner similar to the one in the single viewings. By assuming the line of sight was orthogonal to the screen the relation

$$\text{angle} = \tan^{-1} \frac{\text{screen diagonal}}{\text{distance to screen}} \tag{1}$$

was used to find the maximal angle the eye could move while still viewing the screen. In the single viewings the distance from head to screen varied from 70-90 cm, giving angles of maximal eye movement of 11.2° to 14.3°. The distance from the subject in the centre chair to the screen was measured to be 450 cm and 490 cm for the outermost placed subjects. With a screen diagonal of 102 cm this resulted in angles of maximal eye movement between 11.8° and 12.8°.

The recordings were done in a larger room, to accommodate all the subjects, and the sound from the films was played through loudspeakers, to avoid the emotional distance which noise dampening headphones might produce. On the basis of creating similar lighting conditions as in the single viewing experiment the windows were blacked out and four lamps placed strategically to avoid shining a light in the eyes of the subjects, but still illuminating them for the purpose of filming them. For the joint viewings the subjects were filmed using a GoPro Hero 2. The image and sound quality of the GoPro was not as good as the original camera (especially in the dim lightening), but it had the benefit of being unobtrusive and could be placed directly in front of the subjects. The recording tablets were placed on tables directly behind the subjects to avoid loss of connection from transmitters with poor transmitting distance.

S2.4 Questionnaires and general information about the subjects

Before the EEG recordings all subjects were asked to fill out a questionnaire. Apart from asking relevant physiological questions, it was also chosen to ask the subjects to evaluate their level of proficiency in German, because of the German dialogue in *Sophie's Choice*.

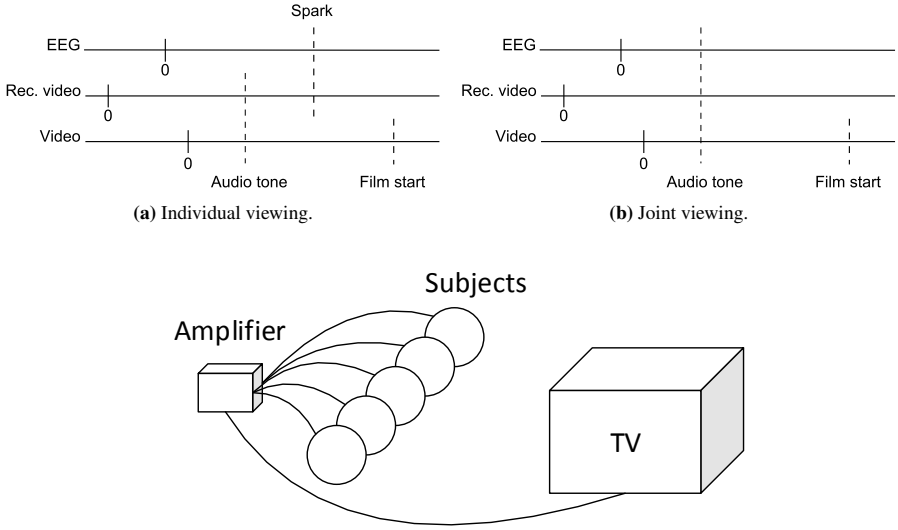
After viewing the films the subjects were asked to answer another questionnaire regarding whether they knew the scenes beforehand and which scenes had the biggest impact on them. Subjects viewing the films with scrambled scenes were also asked to describe the plot in the two films. This was both done to evaluate and possibly subdivide the subjects based on their understanding as well as for a comparison with the results gained from the EEG.

S3 Hardware

Research grade EEG equipment is often very expensive, time-consuming to equip, and immobile. Using smaller consumer grade hardware has thus many advantages if it is able to measure the required signals adequately. Part of the motivation for the experiment presented in this article was to validate if the hardware used is sufficient for the paradigm, in what areas it may be advantageous, and in what areas it is lacking.

S3.1 Emocap

To conduct the experiments the mobile 14 channel consumer EEG headset Emotiv EPOC was rebuild to a wireless cap based on EasyCap, the Emocap. The sampling frequency of the ADC is 2048Hz but since the EPOC only have one ADC the data is



(a) Individual viewing.

(b) Joint viewing.

(c) Joint viewing setup.

Figure S4. Illustration of the synchronisation paradigms for the individual and joint viewing experiments.

sampled sequentially which means that the effective sample frequency of each channel is 128Hz (including Common Mode Reference and Driven-Right-Leg electrodes).[3] Each sample is assigned a number from 0 - 128 in the EPOC in order to ensure detection of packet loss on a sub-second time-scale. The EPOC has previously been validated against the Biosemi Active-II device with 64 channels using an imagined finger tapping paradigm.[4]

The electrode placement of the Emocap follows the 10-20 system in naming and placement, but with only 14 measurement electrodes the configuration is specific to this setup.

S3.2 Tablet

To acquire and record data from the Emocap it is possible to use a computer or a mobile device supporting direct access to the USB port. In this experiment tablets of the model Asus Nexus 7 were used. The processing power of the device is much greater than needed for this application and the tablet has previously been shown to work well with the EPOC.[4]

S3.3 Synchronisation

Experiments involving a stimulus are highly dependent on temporal alignment if the objective is to compare the results across modalities or recordings. To synchronise EEG recordings with the film, two methods were employed based on the experimental condition.

S3.3.1 Individual viewing

For the subjects viewing the films alone the method of synchronisation was based on the theory that the electro magnetic wave, generated by creating a powerful spark, would induce a small current in the wires from the electrodes. This was confirmed using a piezoelectric spark generator normally used to ignite a Bunsen burner. Based on the length of the spark it was estimated that the spark was around 2 kV and very low amperage. When used approximately 2 centimetres from the electrodes a spike with much higher amplitude than the surrounding artefact free EEG was observed. Generation of the spark also emitted a noise that was distinguishable in the audio track of the recorded video. This method was hence used to synchronise the EEG with the

recorded video in single subject experiments. To synchronise the recorded video with the film showed on the tablet, the audio output from the tablet was connected to the input of the camera in parallel with a microphone. At a fixed time before the first film clip a 43Hz sine wave was played to make this part of the synchronisation easier, as illustrated in figure S4a.

S3.3.2 Joint viewing

In the experiments involving simultaneous EEG recordings of multiple subjects, a method of inducing a current simultaneously in all EEG recordings was necessary. Though experiments with creating more powerful sparks with estimated voltages of up to 60 kV, the field from the circuit was not visible in the EEG, meaning that either the signal strength was not adequate or the DRL circuit in the Emocap managed to suppress the signal.

S4 Questionnaires

Table S3. Information regarding the subjects, under which condition they saw the movies, and how they perceived them. Self-reported German proficiency rates is scored from 1 to 4, with 1 being equal to ”None”. **S** = Scrambled scenes, **NS** = Non-scrambled scenes, **J1** = The first joint viewing, **J2** = The second joint viewing.

Subject no.	Condition	Order of movies	Age	Hours of sleep	German proficiency	Right handed	Seen the movies before	Understood the movies
16	S	1	20	7	3	Yes	No	Yes
10	NS	2	21	9	3	No	No	Yes
9	S	1	19	9	2	Yes	No	Yes
15	S	1	22	6	2	Yes	No	Yes
3	NS	2	20	7,5	2	Yes	No	Yes
6	NS	1	20	7,5	3	Yes	No	Yes
11	S	1	18	10	2	Yes	No	Yes
2	NS	5	20	9	3	Yes	No	Yes
4	NS	6	21	7	2	Yes	No	Yes
5	NS	1	21	9	2	Yes	No	Yes
14	S	1	20	9	2	Yes	No	Yes
12	S	1	21	8,5	2	No	No	Yes
13	S	2	20	8,5	1	Yes	No	Yes
7	NS	3	21	7	2	Yes	No	Yes
1	NS	2	19	8	1	Yes	No	Not Sophie's
8	S	2	19	8	1	Yes	No	Not Bang!
23	NS	4	24	7	2	Yes	No	Yes
22	S	1	25	8	2	Yes	No	Not Bang!
24	NS	3	25	8	2	Yes	No	Yes
21	S	1	25	5	2	Yes	No	Yes
17	S	1	21	7	1	Yes	No	Yes
19	S	1	21	7	2	No	No	Yes
18	NS	6	22	7,5	1	No	No	Yes
20	NS	6	23	8	2	Yes	No	Yes
33	J1	6	32	7	3	Yes	No	Yes
28	J1	6	25	8	2	Yes	No	Yes
30	J1	6	24	8	2	Yes	No	Yes
31	J1	6	26	8	1	Yes	No	Yes
27	J1	6	25	5,5	2	Yes	No	Yes
25	J1	6	21	8	2	Yes	No	Yes
29	J1	6	25	10	2	Yes	No	Maybe
26	J1	6	21	8	1	Yes	No	Yes
32	J1	6	25	7	2	Yes	No	Yes
34	J2	6	22	10	2	Yes	No	Yes
36	J2	6	20	9	1	Yes	Sophie's	Yes
35	J2	6	25	8,5	2	Yes	No	Yes
41	J2	6	25	8	2	Yes	No	Yes
37	J2	6	25	9	2	Yes	No	Yes
42	J2	6	22	6	3	Yes	No	Yes
40	J2	6	22	6,5	3	No	No	Yes
39	J2	6	24	8	2	Yes	Sophie's	Yes
38	J2	6	23	8	1	Yes	No	Yes

Questionnaire before EEG-measurement

Name: _____ **subject no.:** _____

Movie: _____

1. Are you right handed? Yes ☐ No ☐
2. Normal sight/corrected to normal vision Yes ☐ No ☐
3. Normal hearing: Yes ☐ No ☐
4. How many hours have you slept last night? _____
5. Age _____
6. Do you have a psychiatric record? Yes ☐ No ☐
7. Do you have a neurologic record? Yes ☐ No ☐
8. Have you ingested drugs or medication the last 24 hours? Yes ☐ No ☐
   a. If yes, which? _____
9. Level of German proficiency
   a. Fluent ☐
   b. Good understanding of the language ☐
   c. Basic understanding of the language ☐
   d. None ☐
10. Are you interested in participating in future experiments? If you answer "Yes", your sex, age and contact information will be saved for future use. Yes ☐ No ☐
11. Can we use pictures of the experimental setup, where you appear for our Master thesis, article or other things that regard this experiment? Yes ☐ No ☐

Mobiles and other electric equipment have to be removed before the experiment.

I hearby confirm, that I agree to participate in a experiment with EEG recordings during viewing films. I am informed that I participate voluntarily and that I can, at any time and without reasons, can redraw my consent to participate.

Date : _____ Signature : _____

Questionnaire after EEG-measurement

Name:		**Subject no.:**	

1. Had you seen the movies before
   a. The black/white movie Yes ☐ No ☐
   b. The movie in colour Yes ☐ No ☐
2. Did you understand the movie in german? (the one in colour) Yes ☐ No ☐

3. Which scenes made the strongest impression in the black/white movie?

_____

_____

_____

_____

4. Which scenes made the strongest impression in the movie in colour?

_____

_____

_____

_____

References

1. Dmochowski, J. P., Sajda, P., Dias, J. & Parra, L. C. Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? *Frontiers in human neuroscience* **6**, 112 (2012).
2. Raz, G. *et al.* Portraying emotions at their unfolding: a multilayered approach for probing dynamics of neural networks. *NeuroImage* **60**, 1448–61 (2012).
3. Emotiv. EPOC specifications (2012).
4. Stopczynski, A., Stahlhut, C., Larsen, J. E., Petersen, M. K. & Hansen, L. K. The Smartphone Brain Scanner: A Portable Real-Time Neuroimaging System. *PloS one* **9**, e86733 (2014).

APPENDIX B

Microstate EEGlab toolbox: An introductionary guide

Authors: **Andreas Trier Poulsen**, Andreas Pedroni, Nicolas Langer, and Lars Kai Hansen.

Status: Available on bioRxiv

URL: https://www.biorxiv.org/content/early/2018/03/27/289850

Toolbox: https://archive.compute.dtu.dk/files/public/users/atpo/Microstate

Microstate EEGlab toolbox: An introductory guide

Andreas Trier Poulsen[1,*], Andreas Pedroni[2,*], Nicolas Langer[2], and Lars Kai Hansen[1]

[1]Technical University of Denmark, DTU Compute, Kgs. Lyngby, Denmark
[2]University of Zürich, Psychologisches Institut, Methoden der Plastizitätsforschung, Zürich, Switzerland
[*]Corresponding authors: atpo@dtu.dk and andreas.pedroni@uzh.ch

March 27, 2018

Abstract

EEG microstate analysis offers a sparse characterisation of the spatio-temporal features of large-scale brain network activity. However, despite the concept of microstates is straight-forward and offers various quantifications of the EEG signal with a relatively clear neurophysiological interpretation, a few important aspects about the currently applied methods are not readily comprehensible. Here we aim to increase the transparency about the methods to facilitate widespread application and reproducibility of EEG microstate analysis by introducing a new EEGlab toolbox for Matlab. EEGlab and the Microstate toolbox are open source, allowing the user to keep track of all details in every analysis step. The toolbox is specifically designed to facilitate the development of new methods. While the toolbox can be controlled with a graphical user interface (GUI), making it easier for newcomers to take their first steps in exploring the possibilities of microstate analysis, the Matlab framework allows advanced users to create scripts to automatise analysis for multiple subjects to avoid tediously repeating steps for every subject. This manuscript provides an overview of the most commonly applied microstate methods as well as a tutorial consisting of a comprehensive walk-through of the analysis of a small, publicly available dataset.

Contents

1 Introduction

Multichannel electroencephalography (EEG) is used to assess the spatio-temporal dynamics of the electrophysiological activity of the brain. Traditionally, researchers analyse EEG by characterising the temporal waveform morphology and/or frequency distribution of recordings at certain preselected electrodes. Even though this method provides a wealth of insights into the electrophysiology of the human brain at work and at rest, this approach still misses out on large parts of the information in the EEG signal and neglects the multivariate characteristics of the measurements. A different approach, that takes into account the information of all electrodes, is to characterise the EEG signal by the spatial configuration of the electric fields at the scalp that can be conceptualised as a topographical map of electrical potentials. The widely used EEGLAB toolbox (Delorme and Makeig, 2004) has promoted the use of independent component analysis (ICA) for such multivariate modelling.

Microstate analysis is an alternative, and increasingly applied, EEG-representation based on topographic analysis. Microstate analysis has its foundation in the work of Dietrich Lehmann and colleagues (e.g. Lehmann, 1971; Lehmann et al., 1987), who observed that the time series of topographies in ongoing EEG are comprised of a discrete set of a few prototypical topographies that remain stable for around 80 - 120 ms before rapidly transitioning to a different topography. These periods of quasi-stable EEG topography have been called *functional microstates* and the discrete spatial configurations *microstate classes*. Following the argument that different EEG topographies are generated by different configurations of neuronal generators, microstates have been suggested to reflect global

functional states of the brain (Khanna et al., 2015; Michel et al., 2009; Lehmann et al., 1998) and have been found to be tightly coupled with functional Magnetic Resonance Imaging (fMRI) resting state networks (Britz et al., 2010; Yuan et al., 2012; Musso et al., 2010; Van de Ville et al., 2010).

Microstate analysis is increasingly recognised as an innovative method offering straight-forward characterisations of brain-states. Its usefulness in gaining insights into brain functions has been demonstrated in healthy (e.g. Koenig et al., 2002) and clinical populations (e.g. Lehmann et al., 2005). Microstates have been examined when subjects are at rest (e.g. Khanna et al., 2015), as well as both during and in between active tasks (e.g. Pedroni et al., 2017). More recently, microstate analysis has been devised in event-related study designs, examining characteristics of pre-and post-stimulus time-locked microstates (e.g. Schiller et al., 2016; Ott et al., 2011).

The output parameters of microstate analysis (described in sections 2.6 and 3.9) quantify the statistical properties of the microstates. They include parameters summarising the variability in the spatial configuration and strength of the topography of microstate classes, parameters pertinent to the temporal dynamics of microstates such as the average duration of a microstate class or the frequency of occurrence, as well as statistics on the transition probabilities between microstate classes. In addition, parsing the EEG into microstates can be used to select epochs of interest (that correspond to a certain microstate class), which can be further examined using other analysis methods such as time-frequency analysis. Also from a practical point of view, compared to functional magnetic resonance imaging (fMRI), microstates analysis of EEG offers a readily applicable and cost-worthy method to investigate temporally coherent network activity.

The core of microstate analysis is to segment the EEG recordings into microstates using a clustering method. There are currently several cluster methods being used for microstate analysis, such as (modified) K-means clustering (Lloyd, 1982; Pascual-Marqui et al., 1995), Topographic Agglomerative Hierarchical clustering (Murray et al., 2008; Khanna et al., 2014), principal component analysis (Skrandies, 1989), or mixture of Gaussian algorithms, with each method reflecting rather a class of clustering algorithms than a completely defined algorithm.

Microstate analysis could arguably have been even more widely adopted had it not been for the relative lack of transparency regarding the methods applied in the field. So far, established toolboxes for microstate analyses have been implemented in compiled software[1], which therefore not revealing all information of the methods. However, transparency about the exact method is in our view important to grant replicability because, as our experience has shown, even small changes in the settings of clustering algorithms sometimes lead to substantial differences in the outcome of the analysis.

The present work aims to facilitate microstate analysis in Matlab by providing a toolbox that is fully transparent with respect to all the steps of analysis and which allows the integration of any clustering algorithm. The toolbox is made of a set of functions that can be used and modified independently or as an interactive plug-in for the widely used, open-source EEG analysis software EEGLAB (Delorme and Makeig, 2004).

We encourage the reader to give suggestions for improvement or self-developed extensions that may be of interest for a broader community. As the development of clustering algorithms is a rapidly advancing field of research, an open and easily modifiable platform to implement new clustering methods for microstate analysis may also foster new approaches to the idea of EEG microstates. For instance, we see potential in advancing methods to select an adequate number of clusters.

The remainder of this article is structured as follows: Section 2, starts by outlining the general methods to conduct a microstate analysis and then provide detailed explanations about three clustering algorithms (i.e. *K-means*, *modified K-means* and *(Topographic) Atomize and Agglomerate Hierarchical Clustering*) that have been commonly used in microstate analyses. We then outline some of measures of fit that can be used selecting the number of microstates, which is followed by a discussion on the settings of microstate clustering. We conclude the methodological part of this guide with

[1]But see: Koenig et al. (2011) for a Matlab implementation of microstates analyses for ERP data, or Milz (2016) for a Python implementation.

by explaining the back-fitting procedure and with an overview of the output of a microstate analysis; the microstate statistics. Section 3 is dedicated to provide an easy entry point into microstates analysis in the form of a tutorial that uses the GUI of the toolbox along with a Matlab script, allowing to automatise and adapt the analysis to new datasets.

2 Microstate methods

This section contains the theoretical part of the guide, where we will walk through the concepts and differences between the methods employed in microstate analysis. We will cover some of the most common clustering algorithms and their settings, as well as approaches on how to select the best solution and how to evaluate it.

In microstate analysis it is the goal to segment the recorded EEG time samples into microstate classes, so EEG samples that belong to the same class have as similar topographies as possible. The segmentation is commonly done using topographical clustering methods, where each cluster of EEG samples denotes a microstate class. For each cluster, the methods calculate a prototypical topographical map based on all the EEG samples assigned to it. The assumption of the clustering is that EEG samples assigned to the same cluster, all originate from the neural processes underlying the prototype topography of that cluster.

In general there are different mathematical and statistical approaches to cluster data, resulting in different clustering methods. How large a differences there is in the resulting clusters, between the clustering methods, depends on the data of interest, and the underlying assumptions of the methods. For example, where some methods ignore the polarity[2] of the topographies, other methods take it into account and assign EEG samples with similar but opposite polarity to different clusters. In section 2.1 we go through the most common clustering methods used in microstate analysis.

After deciding on a clustering method there is, however, still factors that affect the clustering, such as; *How to measure the similarity between two EEG samples*, *should the cluster assignment of a sample be affected by its temporally neighbouring samples*, and *what is the right amount of clusters (microstate classes)*? And after clustering the data it may be relevant to reduce noise by temporal smoothing of the obtained microstate sequence, before calculating statistics from them.

Another thing to note about clustering is that often there is not one true solution that segments the EEG perfectly. In the same way that it is necessary to decide which amount of microstate clusters gives the best result (see section 2.2), some clustering algorithms can converge on different solutions even when using the same number of clusters. This means that using the same method with the same number of clusters might result in different segmentations, when it is run multiple times. See section 2.3.1 for more information on initialisation of algorithms and using 'multiple restarts' to reduce variability.

In the following sections we will describe the different methodological choices that can be made in microstate analysis, with the intent of increasing the understanding of the effect of these choices and how they work. For a walk-through of an example microstate analysis we refer to the tutorial in section 3.

During our review of the methodological side of microstate literature, we encountered several examples of citation of articles that were irrelevant for the cited method, or even citation of articles featuring a completely different method. We have therefore made an emphasis on referencing the relevant articles during the review of microstate methods, with the intention that this section can be used to look up the proper references for specific methods.

[2] see section 2.1.2 for more information on polarity

Table 1: Definitions of the notation employed in this article and in the toolbox.

Notation	Dimensions	Definition
C	$\mathbb{N}$	Number EEG channels.
N	$\mathbb{N}$	Number of time samples.
K	$\mathbb{N}$	Number of clusters (microstate classes).
$\mathbf{X}$	$\mathbb{R}^{C\times N}$	Recorded EEG.
$\mathbf{x}_n$	$\mathbb{R}^{C}$	The n'th time sample of the recorded EEG.
$\mathbf{A}$	$\mathbb{R}^{C\times K}$	Prototypical maps for K clusters.
$\mathbf{a}_k$	$\mathbb{R}^{C}$	The prototypical map for the k'th microstate cluster.
$\mathbf{l}$	$\mathbb{N}^N \in [1;K]^N$	Labels assigning each EEG sample to a cluster.
l_n	$\mathbb{N} \in [1;K]$	The microstate label of the n'th EEG sample.
$\mathbf{Z}$	$\mathbb{R}^{K\times N}$	Microstate activations/amplitude.
$\mathbf{z}_n$	$\mathbb{R}^{K\times N}$	Microstate activations for the n'th time sample.
d_{kn}	$\mathbb{R}$	Distance between $\mathbf{x}_n$ and microstate k for the n'th time sample.
b_{kn}	$\mathbb{R}$	Amount of samples with label k in a window around the n'th time sample.
$\boldsymbol{\epsilon}_n$	$\mathbb{R}^{C}$	EEG recording noise for the n'th time sample (assumed to be zero-mean i.i.d. Gaussian).

2.1 Clustering algorithms

This section contains a short review of the methods implemented in the Microstate toolbox. The toolbox contains the three most commonly used clustering methods for microstate analysis; *K-means*, *modified K-means* and *Topographic Atomize and Agglomerate Hierarchical Clustering*. The toolbox also contains a subgroup of new experimental methods under *Experimental algorithms*.

To help explain the methods we introduce the notation in table 1. The same notation is used in the Matlab code of the toolbox. Note that in microstate literature it is in general assumed that the EEG has been referenced to the average reference.

2.1.1 K-means

K-means clustering (Lloyd, 1982) is a well-established clustering method and also the most basic method employed in the toolbox. It belongs to one of two main groups of clustering methods named the *partitioning methods*, which generally require that the number of clusters (or microstate classes) are pre-set by the user. These methods start with partitioning the EEG samples into the fixed number of clusters, which they relocate the EEG samples to in iterations, until an optimal cluster assignment have been achieved (Rokach and Maimon, 2005).

K-means clustering starts by defining K cluster centres (also known as prototypes), e.g. by choosing K EEG samples at random. It iterates through two steps: First it assigns each EEG sample to the cluster whose prototype it is most similar with. It then re-calculates cluster prototypes, which is often done by averaging over the newly assigned samples. The algorithm continues iterating over these two steps until a convergence criterion has been reached. Two examples of criteria of convergence could be: Stopping when the change in assignments of EEG samples between iterations are low enough to reach a pre-set threshold or when a fixed number of iterations are reached.

From a probabilistic point of view K-means is consistent with the generative model

$$\mathbf{x}_n = \mathbf{a}_{l_n} + \boldsymbol{\epsilon}_n, \quad \text{for} \quad n = 1...N, \tag{1}$$

where $\mathbf{a}_{l_n}$ signifies the topographical map assigned to nth EEG sample.

K-means clustering can be customised e.g. by changing its way of initialising prototypes or how similarity is measured. The K-means clustering implemented in this toolbox employs Matlab's default built-in function **kmeans.m** (Statistics and Machine Learning toolbox required), that uses the **k-**

means++ algorithm (Arthur and Vassilvitskii, 2007) for cluster prototype initialisation and the squared Euclidean metric to determine similarity.

See e.g. Bishop (2006) or Rokach and Maimon (2005) for a more detailed introduction to K-means clustering.

2.1.2 Modified K-means

The modified K-means introduced in Pascual-Marqui et al. (1995) adds a number of features to clustering.

On a practical level there are mainly two differences compared to conventional K-means. The first is that the topographical maps of the prototypical microstates are polarity invariant. This means that samples with proportional, but opposite, topographical maps (e.g. $\mathbf{a}_k$ and $-\mathbf{a}_k$) are assigned to the same cluster. The second difference is that modified K-means models the activations of the microstates, i.e. models the strength of the microstates for each time point.

At the conceptual level, microstate classes (or clusters) are now being viewed as 'directions' in a multi-dimensional topographical space. The activations would then quantify how far along a microstate-orientation the EEG signal is at a given time point. It makes it easier to imagine this using only two EEG channels, which would draw an asterisk in 2D-space. Here each line is a microstate, and time points with a strong EEG signal would be placed further along its microstate line.

Again, taking a probabilistic view point modified K-means is consistent with the generative model

$$\mathbf{x}_n = \mathbf{A}\mathbf{z}_n + \boldsymbol{\epsilon}_n, \quad \text{for} \quad n = 1...N, \tag{2}$$

with the important constraint that only one microstate can be active for each time point, i.e. all K elements of $\mathbf{z}_n$ are zero except for one, i.e. the model can be written

$$\mathbf{x}_n = \mathbf{a}_{l_n} z_{l_n n} + \boldsymbol{\epsilon}_n, \quad \text{for} \quad n = 1...N. \tag{3}$$

The microstate label of an EEG sample is found as the microstate index, k, that minimises the orthogonal squared Euclidean distance,

$$l_n = \arg\min_k \left\{ d_{kn}^2 \right\} \tag{4}$$

$$d_{kn}^2 = \mathbf{x}_n^\intercal \cdot \mathbf{x}_n - (\mathbf{x}_n^\intercal \cdot \mathbf{a}_k)^2. \tag{5}$$

See appendix A for details on this relation as well as an introduction of a new and optimised iteration scheme, for modified K-means. This new iteration scheme upholds the same model assumptions, but preliminary tests show that it is much faster and even slightly better at representing EEG with microstates compared to the original iteration scheme.

The modified K-means method is also accompanied by a scheme for temporal smoothing microstates label sequences to avoid short segments. See section 2.5 for more information on temporal smoothing.

The seminal publication on modified K-means (Pascual-Marqui et al., 1995) contains additional explanation, motivation and pseudo-code.

2.1.3 Topographic Atomize and Agglomerate Hierarchical Clustering

The Topographic Atomize and Agglomerate Hierarchical Clustering (TAAHC) method belongs to the second of two main groups of clustering methods, *hierarchical clustering*. TAAHC was developed for microstate analysis and is a modified version of its precursor, *atomize and agglomerate hierarchical clustering* (AAHC). The (T)AAHC methods differ in significant ways from traditional hierarchical clustering, and even though the only conceptual difference between TAAHC and AAHC is in how

they measure the quality of their clusters, their resulting algorithms are significantly different from each other. See section 2.1.4 for a deeper description of these differences.

In TAAHC, the user does not have to pre-set the number of clusters. It starts out with all EEG samples having their own cluster and then one cluster is removed at a time. Each iteration of the algorithm consists of finding the "worst" cluster, and then remove (atomise) it and reassign each of its members to the cluster they are most similar with. This process is then continued until there are only two clusters remaining (or a pre-set minimum number of clusters).

The "worst" cluster is defined as the cluster that has the lowest sum of correlations between its members and prototype (Khanna et al., 2014)[3]:

$$\mathrm{CorrSum}(k) = \sum_{n}^{N} \mathrm{Corr}(\mathbf{a}_k, \mathbf{x}_n) = \sum_{n}^{N} \frac{|\mathbf{x}_n \cdot \mathbf{a}_k|}{||\mathbf{x}_n|| \cdot ||\mathbf{a}_k||}, \quad \text{for} \;\; l_n = k, \tag{6}$$

assuming average referenced EEG, i.e. that for each time sample, the channel mean has been subtracted. Note that using the absolute value of the correlation makes TAAHC polarity-invariant[4]. Though the TAAHC is a specialised hierarchical clustering, it can also be seen as a specialised K-means in the way it models the EEG.

By using the correlation to find the worst cluster, the TAAHC is not deterministic (see section 2.1.4) unless an extra step is included in the initialisation of the algorithm. To ensure determinism the clusters are initialised by creating two-sample clusters from the most correlated pairs. In the toolbox this is done by calculating correlations between all sample-pairs, assigning the most correlated pair to the first cluster, and removing the two samples from the sample-pool. This process is then repeated on the remaining sample-pool, until all pairs have been found and the number of clusters has been halved. In the case of an odd number of samples, the leftover sample is assigned its own single-sample cluster.

As a side note on references, Tibshirani and Walther (2005) is often referenced when the (T)AAHC methods are used, but this article actually does not explain how agglomerative hierarchical clustering works. Instead it investigates the relevant question of how to validate the correct number of clusters, when using hierarchical clustering. (T)AAHC is also sometimes erroneously referenced to Pascual-Marqui et al. (1995), which pre-dates the methods and instead introduces the modified K-means algorithm.

For more information on different types of hierarchical clustering we recommend Rokach and Maimon (2005). We suggest using Murray et al. (2008) as a reference for AAHC. Being a new method, the TAAHC unfortunately hasn't, to the knowledge of the authors, been thoroughly introduced in a published article. Though Khanna et al. (2014) introduces TAAHC, its determinism-ensuring initialisation scheme is not covered.

2.1.4 The difference between AAHC and TAAHC

Each word, from right to left, in the *(Topographic) Atomize and Agglomerate Hierarchical Clustering* methods denotes a specialisation of hierarchical clustering. *Agglomerative hierarchical clustering* is also known as bottom-up hierarchical clustering, since it starts from the bottom with single-member clusters (Rokach and Maimon, 2005).

Conventionally, in agglomerative hierarchical clustering the two most similar clusters are merged in each step, meaning that all of their members now belong to the same merged cluster. However, Murray et al. (2008) argues this may lead to a "snowball effect", inflating the size of a few clusters. This effect is deemed unwanted as it might interfere with the detection of short periods of stable topography. The AAHC method is specifically designed to counteract this snowball effect.

[3] This is also described in the Cartool help pages by Denis Brunet.

[4] In the toolbox we have made polarity invariance optional, to enable customisation.

Instead of merging similar clusters, AAHC finds the "worst" cluster, disbands (atomises) it, and assigns each of its members to the cluster they is most similar with. In AAHC the worst cluster is defined as the one that has the smallest contribution to the quality of the clustering, as measured by the sum of global explained variance[5] (GEV) of its members (Murray et al., 2008):

$$\mathrm{GEV}(k) = \sum_{n}^{N} \mathrm{GEV}_n, \quad \text{for} \;\; l_n = k. \tag{7}$$

In other words, what makes AAHC different from standard agglomerative clustering is that the members of the removed cluster can join different clusters. This unfortunately also means that you are not able to straightforwardly create a dendrogram, which, though not vital for selecting how many microstates to use, could be used to analyse the structure of the clustered data.

As mentioned earlier, the only difference between TAAHC and AAHC is in how they measure the quality of the clustering. Where the AAHC uses the GEV, TAAHC uses the sum of correlations as seen in (6). This means that TAAHC does not account for the strength of the maps, but only for the similarity of their topographies. Though their definition of cluster quality is the only conceptual difference, it has the effect that the single-member initialisation of TAAHC becomes stochastic, instead of deterministic like AAHC (see section 2.3.1).

TAAHC is stochastic when it initialises using single-sample clusters since all clusters will have a correlation of 1 with their prototypes (because the prototypes are equal to their single member). All clusters are therefore equally bad (or good) and the first "worst" cluster to atomize has to be chosen at random. The AAHC method uses GEV, that weighs the correlation by the global field power (GFP, see section 2.2.1), to find the worst cluster. It is unlikely that two samples have precisely similar GFP, which makes AAHC deterministic. Because of these differences the AAHC has been made available in the toolbox. To ensure determinism, a special initialisation scheme was introduced for TAAHC, as described in section 2.1.3.

Compared to traditional hierarchical clustering it is also an important difference in how the (T)AAHC methods measure similarity (or determine the worst cluster). In standard hierarchical clustering, it is common to use at either the similarity between the cluster prototypes or use the average of all pairwise similarities between cluster members, making the similarity measures independent of the amount of members a cluster has. However, in (T)AAHC the worst cluster is found by summing the correlation or GEV_n (see (6) and (7)) for each cluster member. This way clusters are "awarded" for having more members, even if they are a bad fit.

So even though AAHC was introduced with the aim of preventing a snowball effect (of large clusters absorbing smaller clusters) it might have inadvertently created a different kind of snowball effect in its method for deciding which cluster to atomise. To our knowledge the effect of these differences have not been investigated in published articles.

2.1.5 Experimental algorithms

We have chosen to collect clustering methods that are not known in the microstate community in a submenu, to avoid cluttering the algorithm selection menu. Currently an unpublished method, *Variational Microstate Analysis*, is implemented in the toolbox, and is selectable under *Experimental algorithms*. This submenu is also intended for future experimental clustering methods for microstate analysis.

2.2 Selecting the number of microstates using measures of fit

Though selecting the number of microstates to use is an important choice, it is not a straightforward choice to make. In many situations there is not a single correct answer, but instead many numbers

[5] See section 2.2 for more information on GEV

of microstate clusters that are able to explain your data well. One of the issues is deciding how to measure and validate how well your clusters explain your data (see e.g. Tibshirani and Walther, 2005).

Measures of fit are used to estimate how well different microstate segmentations explains (or fit) the EEG, used to estimate the prototypes. In microstate analysis one of the common approaches to decide on the amount of microstates clusters, is to calculate four measures of fit and then make a qualitative decision based on these measures and the quality of the topographical maps of the microstates (e.g. do they look physiologically feasible?).

Since the toolbox makes microstate analysis available in Matlab, it is also possible for users to create their own tests for selecting the amount of microstate prototypes. This would also make it possible to use cross-validation[6], which is a way control that your segmentation reflects neural activity related to the condition the EEG was recorded under, instead of unrelated recording noise. Briefly, in cross-validation one divides the EEG into a *training set* (e.g. 90% of the EEG), for estimating the microstate prototypes, and a *test set* to test how well the prototypes fit this unrelated data[7].

Below, we will briefly describe the five measures of fit available in the toolbox and refer to Murray et al. (2008) for further discussion on this topic.

2.2.1 Global explained variance

Global explained variance (GEV) is a measure of how similar each EEG sample is to the microstate prototype it as been assigned to. The higher the GEV the better. It is calculated as;

$$\mathrm{GEV}_n = \frac{\left(\mathrm{Corr}(\mathbf{x}_n, \mathbf{a}_{l_n}) \cdot \mathrm{GFP}_n\right)^2}{\sum_{n'}^{N} \mathrm{GFP}_{n'}^2}, \tag{8}$$

where GFP_n is the global field power, which is calculated as the standard deviation across all electrodes of the EEG for the n'th time sample (Murray et al., 2008). The GEV can be seen as the squared correlation between the EEG sample and its microstate prototype weighted by the EEG sample's fraction of the total squared GFP:

$$\mathrm{GEV}_n = \mathrm{Corr}(\mathbf{x}_n, \mathbf{a}_{l_n})^2 \cdot \frac{\mathrm{GFP}_n^2}{\sum_{n'}^{N} \mathrm{GFP}_{n'}^2}. \tag{9}$$

To calculate the GEV for a given cluster, you sum the GEV of each of its members. To use GEV as a measure of fit for a microstate segmentation, you sum the GEV of all samples included in the segmentation.

2.2.2 Cross-validation criterion

The cross-validation criterion (CV) was introduced in Pascual-Marqui et al. (1995). This measure is related to the residual noise, $\boldsymbol{\epsilon}$, and the goal is therefore to obtain a low value of CV.

$$\mathrm{CV} = \hat{\sigma}^2 \cdot \left(\frac{C-1}{C-K-1}\right)^2, \tag{10}$$

where $\hat{\sigma}^2$ is an estimator of the variance of the residual noise calculated as:

$$\hat{\sigma}^2 = \frac{\sum_n^N \mathbf{x}_n^\intercal \mathbf{x}_n - \left(\mathbf{a}_{l_n}^\intercal \mathbf{x}_n\right)^2}{N(C-1)}. \tag{11}$$

[6]Be careful not to confuse cross-validation with the CV criterion, which is a measure of fit.

[7]We recommend reading up on cross-validation before attempting it, as it is possible to make unintentional mistakes, that breaks the division between training and test set.

2.2.3 Dispersion

Dispersion (W) is a measure of average distance between members of the same cluster. For a microstate segmentation with K clusters, the dispersion, W_K is calculated as the sum of squares between the members of each microstate cluster:

$$S_k = \sum_{n}^{N} \sum_{n'}^{N} ||\mathbf{x}_n - \mathbf{x}_{n'}||^2, \quad \text{for} \quad l_n = k \wedge l_{n'} = k \tag{12}$$

$$W_K = \sum_{k}^{K} \frac{S_k}{2N_k}. \tag{13}$$

W_K can be considered as an error measure, where the lower the better. It usually decreases monotonically when increasing the number of clusters, since the cluster prototypes can then be closer to its members. It is therefore not a good measure of fit by itself since it will usually encourage using as many clusters as possible.

Note that the sum of squares calculated in (12) is not a polarity invariant measure. Since it takes polarity into account, it might not be a suitable measure of fit for polarity-invariant methods such as modified K-means and (T)AAHC.

2.2.4 Krzanowski-Lai criterion

The Krzanowski-Lai (KL) criterion was introduced in Krzanowski and Lai (1988), as a means of selecting how many clusters to use based on the dispersion measure. High values of KL usually indicates an optimal number of clusters.

Though the W_K usually decreases monotonically there often exists an "elbow", where the curve flattens. This elbow signifies that there is a drop in the benefit of adding more clusters, and is therefore often chosen as the optimal number of clusters. The KL criterion is a method for detecting such an elbow by looking at the changes in W_K:

$$\mathrm{DIFF}(K) = (K-1)^{2/C} W_{K-1} - K^{2/C} W_K \tag{14}$$

$$\mathrm{KL}(K) = \left| \frac{\mathrm{DIFF}(K)}{\mathrm{DIFF}(K+1)} \right|. \tag{15}$$

The KL criterion attains large values when an elbow in the W_K curve occurs.

In our toolbox we have added the rule that $\mathrm{KL}(K)$ is set to zero when $W_K - W_{K-1}$ is positive, since this indicates that W_K increased instead of decreasing.

Please note, that since the KL is based on W_K it is also not a polarity invariant measure and might not be a suitable measure of fit for polarity-invariant methods such as modified K-means and (T)AAHC. However, since the KL is "just" a method for finding the elbow in the W_K curve, it could be altered to find the elbow in the curve of the measure optimised for by the chosen algorithm. E.g. the modified K-means seeks to optimise the CV criterion, AAHC optimises for GEV and TAAHC optimises for CorrSum(k), (6). This feature has not been implemented in the toolbox, but due to the open nature of EEGlab and our an effort to document the code, it should be possible for the user to customise the toolbox.

2.2.5 Normalised Krzanowski-Lai criterion

The normalised Krzanowski-Lai ($\mathrm{KL}_{\mathrm{nrm}}$) criterion was introduced in Murray et al. (2008) as an adaptation of the original formula, with the intent of computing a normalised curvature of W. It is

defined as:

$$D_K = K^{2/C} W_K - (K+1)^{2/C} W_{K+1} \tag{16}$$

$$\mathrm{KL}_{\mathrm{nrm}}(K) = \frac{D_{K-1} - D_K}{(K-1)^{2/C} W_{K-1}}. \tag{17}$$

In order to only focus on the convex part of the W_K curve, $\mathrm{KL}_{\mathrm{nrm}}$ is set to zero when $D_{K-1} < 0$ or $D_{K-1} < d_K$.

The main difference between the two KL measures is that the original is a *ratio* of change in the dispersion function, where the normalised version look at the *difference* in the change of dispersion. This means that the two measures are not linearly correlated and that they react differently to dispersion curves. Both measures are therefore made available in the toolbox.

Please note that, like the KL, the $\mathrm{KL}_{\mathrm{nrm}}$ is also not a polarity invariant measure and might not be a suitable measure of fit for polarity-invariant methods such as modified K-means and (T)AAHC.

As a side note on references, Tibshirani and Walther (2005) is often referenced when the KL criterion is used, however this article focuses on using new methods to asses the number of clusters instead of using the KL criterion. We suggest referencing Murray et al. (2008) when using the $\mathrm{KL}_{\mathrm{nrm}}$ criterion[8], and the original article, Krzanowski and Lai (1988), when using the KL criterion.

2.3 Settings for microstate clustering

Most clustering algorithms has some settings that influence how they perform in segmenting data. In this section we will explain what some of the important settings of clustering algorithms do, and why they are relevant.

2.3.1 Multiple restarts of algorithms

In the introduction of this methods section it was mentioned that some algorithms gives a different result every time they are run, even with the same settings. To explain the reason for this it is necessary to make a distinction between *deterministic* and *stochastic* algorithms.

As the name implies, a deterministic algorithm is an algorithm that always give the same result, when its settings and the data stay the same. In other words there is no randomness in the algorithm. If there is randomness somewhere in the algorithm, it is stochastic. This randomness is often introduced in how the algorithm is initialised.

When using K-means or modified K-means for microstate analysis, the cluster prototypes are often initialised by selecting EEG samples. There is usually not one true way of selecting which EEG samples to use for prototypes, so a common approach is to select the EEG samples at random, making the algorithm stochastic.

Though it would be tempting to always select the EEG samples in the same manner to ensure determinism, it runs the risk of obtaining an inferior result. By restarting the stochastic algorithm multiple times, you are able to test multiple segmentations on the same dataset and select the best restart based on e.g. GEV. The downside is that you are not guaranteed to obtain the same segmentation every time the algorithm is run.

How many restarts to use is a trade-of between computation time, and how likely you are to converge on the same optimal solution. In the toolbox 10 restarts are used by default.

2.3.2 Stopping criteria: Convergence and maximum iterations

Some algorithms, like K-means, work by repeating multiple steps in iterations, which means that it will keep iterating until some stopping criteria is met. One of these criteria is the convergence

[8]The $\mathrm{KL}_{\mathrm{nrm}}$ is the criterion implemented in stand-alone microstate software, CARTOOL.

threshold, which stops the algorithm when the relative change in error between iterations drops below the threshold. By default this is set to 10^{-6} in the toolbox.

Again there is a trade-of between computation time and precision of the segmentation, for which reason it is also possible to set a maximum number of iterations before stopping. By default the toolbox has a maximum number of iterations set to 1000.

2.4 Backfitting microstates to new EEG

When you have obtained microstate prototypes from your chosen clustering method, you might wish to see how well some other recordings fit with these prototypes. This is relevant, e.g., when working with spontaneous recordings, where you can't reduce the noise levels through averaging trials. Instead you only give the clustering algorithm time samples that peaked in the GFP time curve, as they are assumed to have the "cleanest" representations of their microstate. After obtaining the prototypes it is then relevant to see which microstate all prototypes likely belong to.

Backfitting is assigning a microstate labels to EEG samples based on which microstate prototype they are most topographically similar with. This similarity can be measured using global map dissimilarity (GMD).

The GMD, which is also known as DISS (Murray et al., 2008), is a distance measure that is invariant to the strength of the signal and instead only looks at how similar the topographical maps look. For two EEG samples, $\mathbf{x}_n$ and $\mathbf{x}_{n'}$, GMD is calculated as:

$$\mathrm{GMD} = \frac{\left|\left| \frac{\mathbf{x}_n}{\mathrm{GFP}_n} - \frac{\mathbf{x}_{n'}}{\mathrm{GFP}_{n'}} \right|\right|}{\sqrt{C}}. \tag{18}$$

By normalising with GFP, two EEG samples that belong to the same microstate, but have different strength, will achieve a low GMD distance.[9]

2.5 Temporal smoothing

EEG recordings are known to contain a lot of unwanted noise, especially in spontaneous recordings where the noise can't be averaged out like with ERPs. This noise can contribute to short (i.e. few or single-sample) microstate segments appearing after clustering or backfitting. One way to address this is to use temporal smoothing, where an EEG sample is not only assigned to a microstate class based on topographical similarity with the prototype, but also based on the microstate labels of samples prior and following the EEG sample.

The most commonly used clustering algorithms in microstate analysis do not take the temporal order of the EEG samples into account. Therefore the temporal smoothing will be done as a processing step after the microstate segmentation has been run.

Currently there are two variations of post-segmentation smoothing implemented in the toolbox; *windowed smoothing* and *small maps rejection*.

2.5.1 Windowed smoothing

This variation of temporal smoothing of microstate segments was introduced in Pascual-Marqui et al. (1995).

The microstate segments are smoothed by updating the microstate labels using a distance measure, that is a trade-off between optimising the best fit to data and temporal smoothness of the labels.

[9] Note that the article Lehmann and Skrandies (1980) is often cited for introducing GMD, and while it does introduce the concept of measuring topographical similarity between two microstate time samples, it does not introduce the equation. See e.g. Michel et al. (2009); Murray et al. (2008) for the equation and more information on GMD.

This distance measure is obtained by adding a bonus term for temporal smoothness to the distance measure expressed in (5);

$$d_{kn}^2 = \mathbf{x}_n^\intercal \cdot \mathbf{x}_n - (\mathbf{x}_n^\intercal \cdot \mathbf{a}_k)^2 - \lambda \cdot b_{kn}, \tag{19}$$

where b_{kn} is equal to the amount of samples that has the microstate label k in a window surrounding the n'th sample, and λ denotes how strongly to weigh smoothness.

This method of smoothing needs the user to set two parameters; λ and the width of the window used. Pascual-Marqui et al. (1995) suggets to set λ to 5, and the window width to three samples (on each side of the current sample), however, the optimal settings might vary between datasets, e.g., due to varying sampling frequency or recording conditions. To find the optimal smoothing settings it would therefore be necessary to use cross-validation.

2.5.2 Small segment rejection

For this toolbox, we introduce a method which sets a minimum duration microstate segments are allowed to last. The algorithm repeatedly scans through the microstate segments and changes the label of time frames in such small segments to the next most likely microstate class, as measured by GMD. This is done until no microstate segment is smaller than the set threshold.

2.6 EEG microstate statistics

Parsing the EEG data into microstates (i.e. labelling of the EEG data with respect to the best fitting microstates classes) offers a rich set of statistical parameters with potential neurophysiological relevance. These statistical parameters can be divided into parameters about the activation strength, the spatial configuration and the temporal attributes of microstates. The strength of the average global activation during a given microstate k is defined by its *average* GFP of all EEG samples assigned to microstate k:

$$\mathrm{GFP}_k = \frac{1}{N_k} \sum_n^N \mathrm{GFP}_\mathrm{n}, \quad \text{for} \;\; l_n = k, \tag{20}$$

where N_k is the number of samples assigned to cluster k. See section 2.2.1 for more information regarding GFP.

The *average* GEV (see equation 8 and 9) and the *average spatial correlation* (between microstate prototype maps and their assigned EEG samples) reflect to what extent the microstate protoypes can explain the data. Where GEV looks at both the strength of the EEG and the spatial fit, the average spatial correlation only looks at the spatial fit.

$$\mathrm{GEV}_k = \frac{1}{N_k} \sum_n^N \mathrm{GEV}_n, \quad \text{for} \;\; l_n = k \tag{21}$$

$$\mathrm{Corr}_k = \frac{1}{N_k} \sum_n^N \mathrm{Corr}(\mathbf{a}_k, \mathbf{x}_n), \quad \text{for} \;\; l_n = k \tag{22}$$

$$= \frac{\mathrm{CorrSum}(k)}{N_k} \tag{23}$$

Note that in the toolbox the correlation is polarity-invariant, i.e. non-negative.

The basic temporal dynamics of microstates are described by Occurrence(k), Duration(k), and Coverage(k). In the toolbox these are calculated using run-length encoding. Occurrence(k) reflects the average number of times per second a microstate is dominant, the Duration(k) is defined as the

average duration of a given microstate (in milliseconds), and the Coverage(k) reflects the fraction of time a given microstate is active.

In addition to these basic statistical parameters, the *transition probabilities* between microstates can be derived to quantify how frequently microstates of a certain class are followed by microstates of other classes. There are (at least) two variants of this statistic, one where the transition probabilities are not corrected for different number of occurrences of microstates (i.e. base rates of microstates) and one which adjusts for the base rate probabilities. Note that only the former version is currently implemented in the toolbox.

3 Guide to the Toolbox

In this part, we provide a practical guide how to conduct EEG microstate analyses using the EEGlab toolbox. Figure 1 provides an overview of the specific processing steps pertinent to the analysis of spontaneous EEG data and for event-related potential (ERP) data with the corresponding `Matlab functions` and typical arguments that have to be passed to the functions. In the subsequent sections we will provide a more thorough tutorial that shows how to analyse a small dataset of spontaneous EEG data (see 3.1). We would like to emphasise that our choices regarding algorithms, settings, and even the whole work-flow itself are based on our experience and not on empirical evidence, which is, to the best of our knowledge, lacking so far.

3.1 Tutorial: Microstate Analysis on spontaneous EEG data

This tutorial will demonstrate how to use the graphical user interface (GUI) of the Microstate Toolbox for EEGLAB to perform a microstate analysis on a small set of spontaneous EEG data with four subjects. The data used for this tutorial can be downloaded here: `https://archive.compute.dtu.dk/files/public/users/atpo/Microstate` ($\sim$ 350 Mb). The archive also contains the newest version of the toolbox and this guide. With this dataset, it should be possible to reproduce the actions discussed in the tutorial and to obtain similar results as shown in the figures and outputs.

To be able to use the Microstates Toolbox presented in this tutorial, Matlab and EEGLAB (Delorme and Makeig, 2004) needs to be installed and the tutorial data should be stored in an accessible folder. To install the toolbox, simply unzip the toolbox into your MATLAB/eeglab/plugins folder. The toolbox is tested with Matlab version 2015b or later and EEGLAB version 14.0.0 or later.

3.1.1 Description of the tutorial dataset and completed preprocessing steps

The tutorial data consists of EEG from 4 subjects (with the original IDs: A00062279001, A00062578001, A00062842001, A00062219001) of the age group 25-44 from the data sharing paper by Langer et al. (2017)[10]. The datasets have been preprocessed using the Automagic[11] artefact correction (with the standard settings) and have been filtered with a highpass filter of 1 Hz and a lowpass filter of 30 Hz using the standard settings of pop_firfiltnew() from EEGLAB 14.0.0b.

3.1.2 Automatising analysis steps for multiple datasets

After finishing this tutorial, the reader may imagine that conducting the analysis steps using the GUI for larger datasets may become a tedious and an error-prone endeavour. A more efficient and reproducible way of conducting the same analyses can be accomplished using a MATLAB script that calls the functions that perform the specific processing steps for all datasets as a batch process. Such

[10] The original, raw-data is available at `http://fcon_1000.projects.nitrc.org/indi/cmi_eeg/eeg.html`
[11] Available at `https://github.com/amirrezaw/automagic`

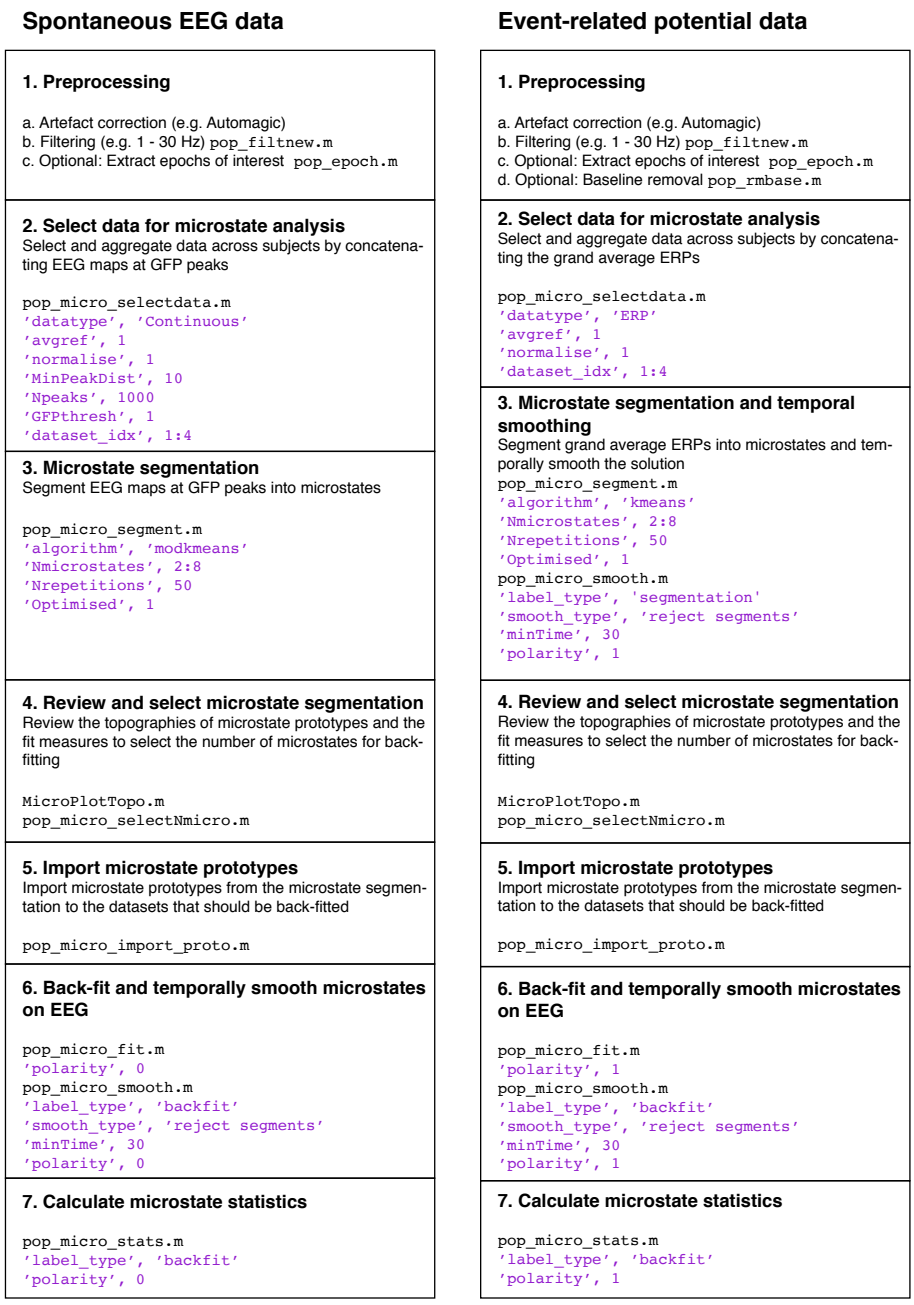


Figure 1: Schematic overview of a workflow for spontaneous EEG and ERP data.

scripts can be easily compiled using EEGLAB's history command `eegh`[12]. Furthermore, a script that

[12] see https://sccn.ucsd.edu/wiki/Chapter_02:_Writing_EEGLAB_Scripts for a description on using this command

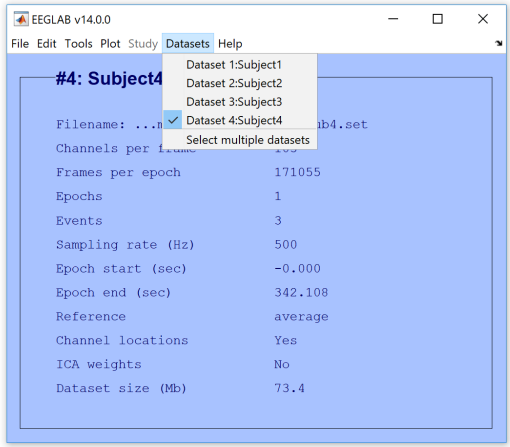


Figure 2: Datasets loaded into EEGLAB.

performs the exact analysis steps of this tutorial is provided in appendix B.

3.2 Outline

To recap, the goal of a microstate analysis is to parse EEG maps into microstate prototypes and re-express the spatio-temporal characteristics of the EEG time series by means of the microstate prototypes. Thus, the first step (after loading the data) is to derive a set of microstate prototypes by clustering the EEG data into as few as possible microstate prototypes that account for as much as possible variance in the EEG data. For the microstate analysis of spontaneous EEG, such as resting state data (as opposed to time-locked EEG data), the amount of data is typically reduced by only using the EEG maps, where the GFP is peaking. This reduces the number of EEG maps that enter the analysis (and hence the processing time) and discards maps of low signal-to-noise ratio. The selected EEG at GFP peaks enters a clustering algorithm that groups it into a small set of classes based on topographic similarity and calculates for each microstate class a topographical prototype, the so called microstate prototype.[13] After obtaining the microstate prototypes, the topography of each EEG sample is labelled as belonging to one of these classes, and the EEG signal is re-expressed as a sequence of microstate classes. Finally, statistics about the sequence of microstate classes, such as their frequency of occurrence or average duration can be calculated. These statistics are typically submitted to further statistical analyses such as group comparisons.

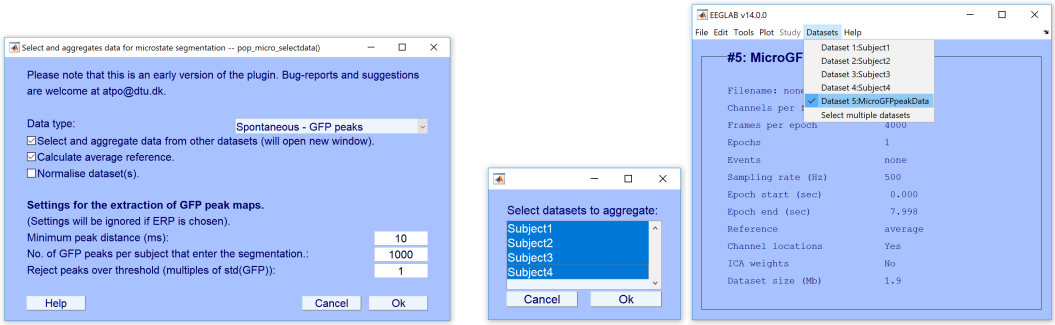
3.3 Data selection and aggregation

3.3.1 Loading datasets in EEGLAB

Start EEGLAB and load each of the four datasets with `Load existing dataset`[14]. This step has to be repeated for each dataset. All loaded datasets can be viewed and selected from the `Datasets` menu (see Figure 2).

[13] Given that the selected EEG maps at GFP peaks are temporally independent, the clustering algorithm ignores information about the order of their appearance, which is the reason that we do not apply temporal smoothing to the segmentation, in contrast to the microstate segmentation of event-related EEG.

[14] See https://sccn.ucsd.edu/wiki/Chapter_01:_Loading_Data_in_EEGLAB for details on installing and starting.



(a) Settings for selecting data for segmentation.

(b) Aggregate from multiple datasets.

Figure 3: Select your settings for the data to be segmented. If you are aggregating from multiple datasets, another window will prompt you to select which datasets to aggregate from, and a new dataset will be created.

3.3.2 Select data for microstate segmentation

Before starting the microstate segmentation, it is necessary to specify the data of which the microstates will be derived from. Since, we are typically interested in obtaining microstate prototypes that explain variance in datasets consisting of more than one subject, the toolbox provides functions to aggregate EEG from multiple datasets into one dataset. This dataset is then used to derive microstate prototypes from.

To select data, open the menu `Tools/Microstates/Select data for microstate analysis`. This will open a window (Figure 3a), which allows you to select the following options (with suggested settings):

- *Data type*: By selecting *Spontaneous - GFP peaks*, EEG maps that represent peaks in the GFP time curve will be used for segmentation. By selecting *ERP*, EEG from all time points will be used. If the dataset contains multiple epochs/trials, these will be averaged into one average epoch (i.e. a grand average ERP). Since the data used for this tutorial stems from spontaneous EEG, we set the data type to *Spontaneous*.

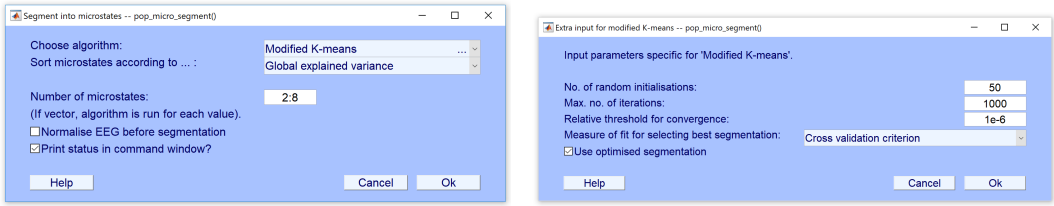
- *Select and aggregate data from other datasets*: This option allows to perform the microstate segmentation on EEG from multiple datasets. In the case of *Spontaneous* data, for each subject, a selection of defined number of EEG maps at GFP peaks is randomly extracted and concatenated to one new dataset[15]. This option will open a new window, where the datasets that are already loaded into EEGLAB are listed and can be selected. Note, if this option is not selected, the segmentation is only performed on the currently active dataset.

- *Calculate average reference*: We advise to reference the data to average reference. Michel et al. (2009) discuss why it is a good idea to use the average reference.

- *Normalise dataset(s)*: This option will normalise each dataset with by the average channel standard deviation within each dataset. This makes the amplitude of the EEG comparable between recordings, which might help weighing each dataset equally in the microstate clustering. However, if the data contains high amplitude artefacts the normalisation might skew the weight given to each dataset. For this walk-through we will not select this option.

- Settings for the extraction of GFP peak maps - These are only relevant and active when selecting *spontaneous* data:

[15] In the case *ERP* has been selected, the epochs/ERPs from each dataset are either averaged into one grand average ERP or are concatenated in time (resulting in one long epoch).



(a) Select clustering algorithm. **(b)** Settings for modified K-means.

Figure 4: Settings for microstate segmentation. After selecting a clustering algorithm, another window will appear with settings specific to the selected algorithm.

- *Minimum peak distance (ms)*: Defines what the minimal distance between GFP peaks can be. This is to ensure only using distinct peaks. We set the minimum peak distance to 10 ms.
- *No. of GFP peaks per subject that enter the segmentation*: Indicates how many GFP peaks are maximally extracted per dataset. We choose to extract EEG samples from 1000 randomly selected GFP peaks for each subject.
- *Reject peaks over threshold*: Discards maps with a GFP that exceeds X times the standard deviation of the GFPs of all maps. This option exists because maps with extreme GFPs often include artefacts of non-neural origin with high amplitudes. Here we will reject maps with GFP that exceeds 1 times the standard deviation of GFPs. Note that this rather low threshold prevents from an unwanted influence of outliers that can occur when analysing data from only four subjects. If there are few GFP peaks available per subject, it may be reasonable to select higher values for this setting.

Confirm the settings by clicking *Ok*. A new window *Select datasets to aggregate* (see Figure 3b) appears to select the datasets to include in the analysis. Use the Shift or Ctrl keys to select multiple datasets and confirm your selection by clicking *Ok*. A new dataset will be created called *MicroGFPpeakData* (which is now listed under `Datasets`) that includes the EEG extracted and aggregated based on the GFP peaks.

Please note, that we are going to estimate microstate prototypes based on GFP peaks selected at random, and there will therefore likely be slight differences in the resulting microstates compared to the ones illustrated on figures 5b and 9b. We experienced converging results when using 1000 peaks per subject for this specific dataset. However, note that for other datasets it may be necessary to take more GFP peaks.

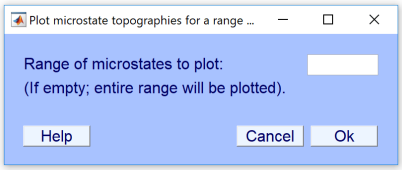
3.4 Microstate segmentation

In this step, the selected EEG (in our case aggregated GFP peaks from four subjects found in *Datasets: MicroGFPpeakData*) is segmented into a predefined number of microstate prototypes with the goal of maximising the similarity between the EEG samples and the prototypes of the microstates they are assigned to. How this similarity is measured is dependent on the chosen clustering algorithm (see section 2.1).

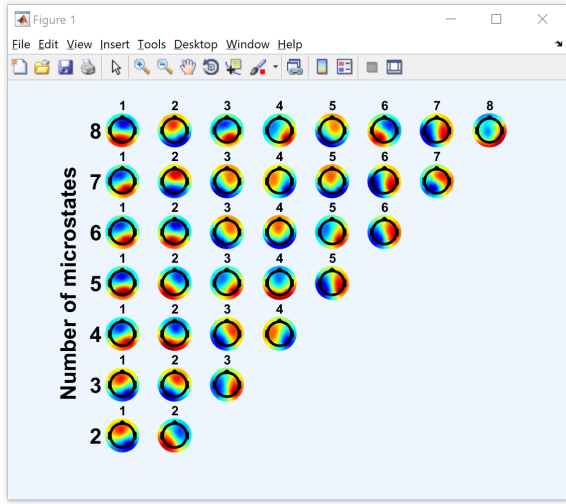
Make sure that the dataset *MicroGFPpeakData* is the currently active dataset (indicated by a tick in the menu `Datasets`) before opening the menu `Tools/Microstate analysis/Segment into microstates` to select (Figure 4a).

As we are working with spontaneous EEG data, we will choose the *modified K-means* algorithm[16],

[16]The (T)AAHC algorithms are also polarity invariant and could be used. The K-means algorithm implemented in this toolbox, however, accounts for polarity because it uses the Euclidean distance as a similarity measure. K-means is



(a) Select which segmentations to plot.



(b) Prototype topographies from 2 to 8 clusters.

Figure 5: A qualitative assessment of the prototype topographies can be used to decide how many clusters to use. Note that since modfied K-means is polarity-invariant the prototypes for different segmentations might be inverted as is the case for prototype number 1 for 2 and 3 clusters.

since it ignores the polarity of the EEG topography (Pascual-Marqui et al., 1995; Michel et al., 2009; Wackermann et al., 1993; Lehmann, 1971).

By default the microstate prototypes are sorted by decreasing *Global explained variance*. The sorting is done after the clustering and does not affect the segmentation. We set the sorting to its default setting and set the *Number of microstates* to cluster the EEG data into from 2 to 8 microstate prototypes (which seems reasonable considering that a majority of studies find four microstate prototypes to be most adequate to describe resting EEG data).

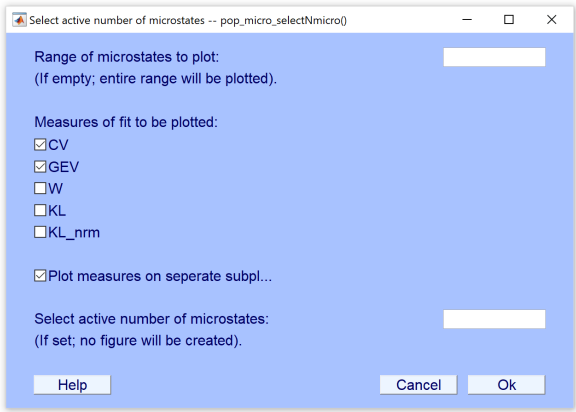
It is possible to *Normalise EEG before segmentation* by the average channel standard deviation, as was the case in `Select data for microstate analysis`. However, here the normalisation will be done across the entire dataset and only for the segmentation, i.e. the stored data will not be changed. This could be relevant, if one wanted to keep the dynamic ranges between datasets by not normalising during data aggregation, but still wanted to help the clustering methods that sometimes assume unit standard deviation. For the same reasons as earlier, we will leave the option unchecked for this walk-through.

After confirming the settings with *Ok* an additional window will open (Figure 4b), in which we can set parameters for the modified K-means algorithm. As mentioned in section 2.3 the first three settings reflect a trade-off of reproducibility versus computation time.

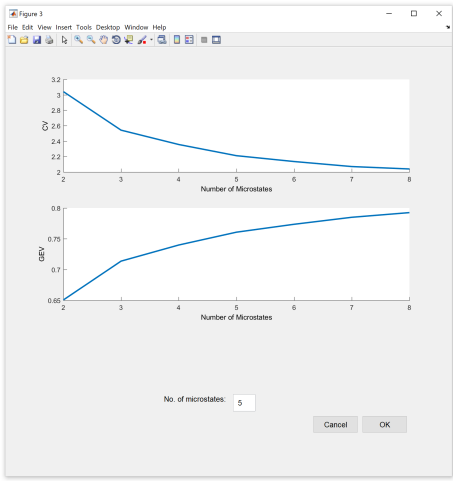
In general, it is advisable to set as high an amount of *No. of random initializations* as computationally feasible to obtain a replicable result. This can be tested by repeatedly running the segmentation and comparing the topographies of the resulting microstate prototypes. If the algorithm reaches the pre-set *Maximum number of iterations* for all restarts (as can be observed in the output on the command window in Matlab), it is recommended to increase this value. The *Measure of fit for selecting best segmentation* defines how to decide which of the N restarts obtained the best clustering. As modified K-means seeks to optimise the clustering based on the CV criterion, we will also set this as the measure of fit.

We *Use optimised segmentation* to employ the optimised iteration scheme for modified K-means

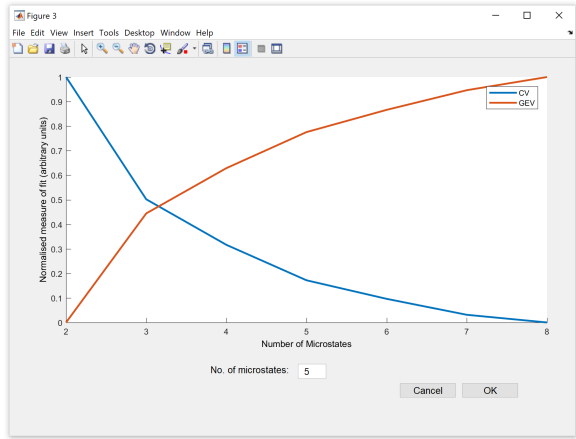
therefore better suited for ERP analysis, where one might want to take polarity into account.



(a) Select which measures of fit to plot.



(b) Separate subplots.



(c) Using same figure.

Figure 6: Select active number of microstates based on measure of fit. These can either be plotted on individual subplots or in the same figure. If plotted together, the plotted measures of fit are normalised to lie in the range $[0; 1]$. When clicking *Ok*, the segmentation from with the *No. of microstates* will become the active segmentation.

that is introduced in appendix A. We use this new iteration scheme, since our preliminary tests indicate that it is much faster and is slightly better at representing the EEG as microstates. The excat computational speedup depends on the dimensionality of the segmented data, but for the present data the optimised iteration scheme results in a computational speed-up by a factor of $\sim$5. We will use this speed-up to increase the *Maximum number of iterations* to 50 and leave the other two settings to their default values.

3.5 Review and select microstate segmentation

After having clustered the EEG with multiple numbers of microstate clusters it is necessary to select which number of clusters to use for further analysis. In microstate analysis, this is often decided based on a qualitative evaluation of the prototype topographies and measures of fit. In some study that analyse spontaneous EEG, while participants are at rest, 4 clusters are chosen, based on previous

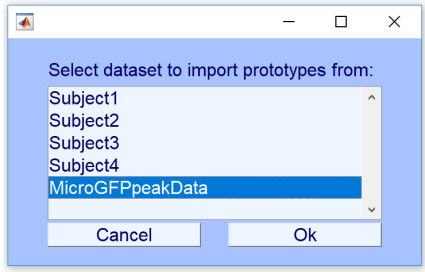


Figure 7: Import microstate prototypes from the *MicroGFPpeakData* dataset into the current dataset.

results, summarized in (Michel and Koenig, 2017).

3.5.1 Plot microstate prototype topographies

To review the result of the microstate segmentations for different numbers of microstates, we look at the quality of the microstate prototypes topographies by selecting `Plot microstate prototype topographies`. This opens a figure as seen on Figure 5b that shows the microstate prototype topographies of each microstate segmentation, sorted by the measure selected in `Segment into microstates` from left to right. The qualitative assessment of the microstate prototypes can, e.g. be based on if the same microstate segmentation contains prototypes with very similar topography or whether the topographies seem physiologically feasible.

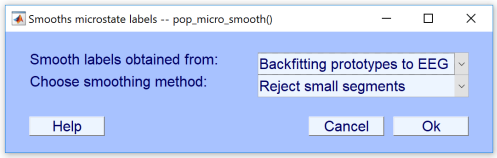
3.5.2 Select active number of microstates

In addition to selecting the number of microstates based on the topographies, we should review the quality of the different microstate segmentations with respect to the goodness of fit. There are various measures for this as explained in section 2.2. Since these measures are calculated in different ways, segmentations that score well with one measure do not necessarily score just as well with other measures. Importantly, the fit measures W, KL and KLnorm are not polarity-invariant, as it is assumed in the segmentation of spontaneous EEG data. Therefore, we should only use the GEV and the CV criterion to review the goodness of fit of our microstate segmentations. To do this, open the menu `Tools/Microstate Analysis/Select active number of microstates` (Figure 6a) and deselect W, KL and KLnorm and only select GEV and CV and click *Ok*.

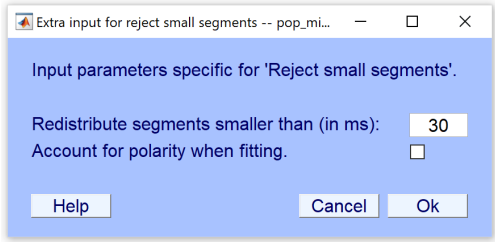
A new figure shows the measures of fit plotted for the different microstate segmentations. This is either plotted on subplots (Figure 6b) or normalised together (Figure 6c), depending whether we have previously selected *Plot measures on separate subplot*. If plotted together, each measure is normalised independently to lie between 0 and 1.

The decision for the right number of clusters obviously reflects a trade-off between the goodness of fit and the complexity a high number of microstates brings to the segmentation. The GEV criterion theoretically (and most of the time effectively) becomes monotonically larger, when increasing the number of clusters, and it is therefore a question of stopping when adding another cluster does not bring a significant benefit.

Note that it can happen that the GEV does not monotonically increase. This indicates that the segmentation has not arrived at the best clustering solution, which might be overcome by increasing the number of restarts and maximal number of iterations and decreasing the threshold for convergence to ensure the best solution is obtained for each number of microstates. while the GEV does not account for complexity (i.e. degrees of freedom), the CV criterion does. However practically, the CV criterion, pointing to the best clustering solution at its smallest value, often reaches such a minimum only with large numbers of clusters, which are difficult to interpret.



(a) Select which labels to smooth and how.



(b) Settings specific for small segments rejection.

Figure 8: Temporal smoothing microstate labels. The microstate labels can either stem from prototypes backfitted to EEG, or from the EEG used in the microstate clustering.

In our case an argument could be made for choosing between four and six clusters. We select using five clusters since it has a good balance between goodness of fit and the number of clusters, as well as it obtains feasible topographies.

3.6 Back-fit microstates on EEG

Now that we have selected the number of microstate prototypes, we are ready to fit the microstate prototypes back to all of the EEG (and not only the GFP-peak data we selected for segmentation). In this processing step, we essentially label the EEG data with the class of the microstate prototype that is most similar. To do this we need to import the microstate prototypes to the dataset to which the prototypes are backfitted to.

We start by changing the active dataset in `Datasets` to the first dataset (Subject1). We want to import the microstate prototype topographies we just obtained, which is stored in the dataset *MicroGFPpeakData*. We therefore select `Import microstate prototypes from other dataset` and then *MicroGFPpeakData* from the list of datasets (see Figure 7). This will add the field `microstates.prototypes` to the `EEG` structure in the Matlab workspace.

As a last step in the back-fitting, we select `Backfit microstates on EEG` to open a window, in which we can choose whether we want to ignore polarity when back-fitting. As we are analysing spontaneous EEG, we want to ignore polarity. So leave this box unchecked and press *Ok*.

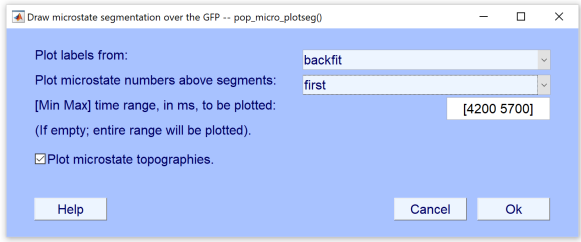
3.7 Temporally smooth microstate labels

In the back-fitting procedure, each EEG sample is assigned to the class of the microstate prototype it is most similar with. As spontaneous EEG includes noise, it will happen often that consecutive time frames are labelled differently by chance. In addition, short periods of unstable EEG topographies occur typically during the change of polarity, even in the same microstate. To reduce these spurious influences, microstate labels can be temporally smoothed after the back-fitting. To do this we open the menu `Temporally smooth microstates labels` (See Figure 8a).

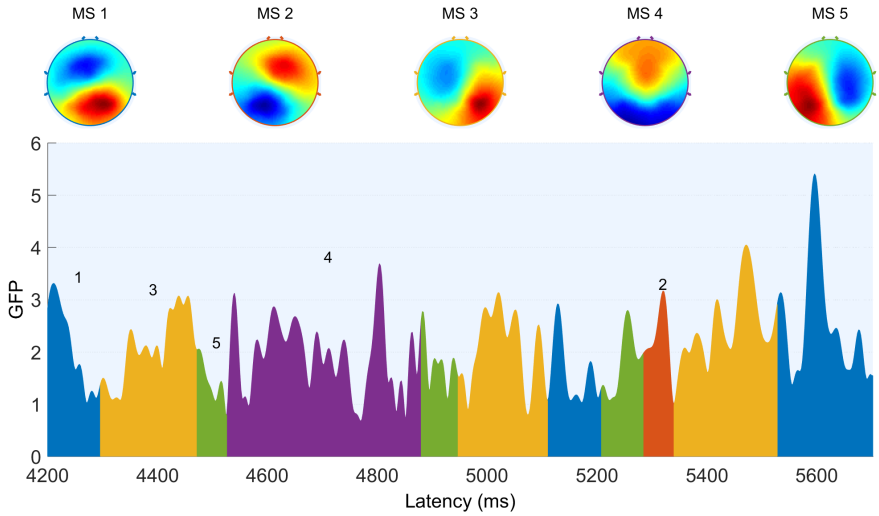
In *Smooth labels obtained from:* select *Backfitting prototypes to EEG*. Here we select the smoothing method *Reject small segments*.[17] Click *Ok*.

A window opens (Figure 8b), where we can indicate how small microstate segments that we will tolerate before the EEG samples are redistributed to other microstate clusters. We use the default settings and click *Ok*.

[17] see section 2.5.1 for more information on the two options for temporally smoothing microstate labels.



(a) Select time range to plot.



(b) GFP of active microstates.

Figure 9: Illustrative figure of the GFP of active microstates for 1,500 ms, from 4,200 ms to 5,700 ms, of the EEG from the first subject.

3.8 Illustrating microstate segmentation

One might wish to visualise the microstate segmentation. This works especially well for ERP analysis, where the entire epoch can fit in a single figure. Since we use spontaneous data for this guide we will instead visualise 1.5 s from subject 1. Figure 9 shows the GFP of the active microstates in the time range between 4,200 and 5,700 ms of the entire resting EEG dataset for subject 1.

3.9 Calculate microstate statistics

Once we have back-fitted to all of the EEG and temporally smoothed the microstate labels, we can calculate the statistical properties of the microstates (see section 2.6), including the average GFP in all time frames of the same microstate class, the occurrence of each microstate class per second, the average duration, the percentage coverage of each microstate class, the Global Explained Variance and spatial correlation (i.e. on average how much variance in the EEG data is explained by the best fitting microstate prototype) and the transition probabilities between microstate classes. To run these calculations click on `Calculate microstate statistics`. This will open a window (Figure 10), in which you select *Use labels obtained from Backfitting prototype* to EEG and leave the entry for *Time window of analysis* empty, so it uses all time frames. Leave the option *Account for polarity when fitting* unchecked, as we have not accounted for polarity previously in the analysis.

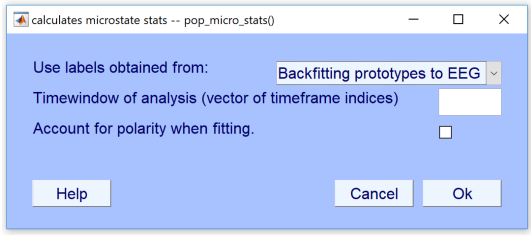


Figure 10: Settings for microstate statistics.

The results of the microstate statistics are stored in the `microstate.stats` field of the `EEG` structure in the Matlab workspace. It is then possible for the user to submit these statistics to further custom analyses.

Repeat all steps in sections 3.6, 3.7 and 3.9 for each dataset. See section 3.1.2 on how to automatise these steps to avoid going through the steps in the GUI for each dataset.

4 Concluding remarks

The aim of both the toolbox and guide has been to make it easier to adapt to microstate analysis for researchers new to the field. Both by bringing transparency to the employed methods, and by supplying an open source toolbox in Matlab.

It is our impression that researchers that are interested in microstate analysis hesitate to adopt to the field, due to a lack of understanding of the underlying methods. With our guide we hope to increase understanding of the methods and hopefully clear up some of the confusion that exits regarding references and miss-citations of articles when citing methods.

By existing in the Matlab framework, the toolbox both makes it easier for users to customise their analysis to suit their experimental conditions and to make the analysis less tedious by being able to create script that can loop the analysis over an entire cohort of subjects and different conditions. We have also sought to thoroughly document the code of the toolbox. Both so users can see how the algorithms are implemented to increase transparency, but also it makes it easier to add new methods to the toolbox in the future.

This manuscript has been intended both as a guide to the toolbox and as an overview of the methods employed in the field of microstates. Therefore, if you have things that you feel are missing or find errors in either guide or the toolbox, please lets us know.

Acknowledgements

We owe thanks to Frans Zdyb for helpful discussions and his preliminary work for a toolbox. We would also like to thank Scott Makeig for helpful discussions and comments for this manuscript. Finally, we thank Denis Brunet for helpful discussions regarding the TAAHC method.

The work was supported by the Novo Nordisk Foundation Interdisciplinary Synergy Program 2014, Biophysically adjusted state-informed cortex stimulation (BASICS) with grant number NNF14OC0011413, and by the Swiss National Science Foundation (100014_175875).

Additional information

Competing financial interests The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Arthur, D. and Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics.

Bishop, C. (2006). *Pattern recognition and machine learning*. springer.

Britz, J., Van De Ville, D., and Michel, C. M. (2010). BOLD correlates of EEG topography reveal rapid resting-state network dynamics. *Neuroimage*, 52(4):1162–1170.

Delorme, A. and Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

Khanna, A., Pascual-Leone, A., and Farzan, F. (2014). Reliability of Resting-State Microstate Features in Electroencephalography. *PLoS ONE*, 9(12):e114163.

Khanna, A., Pascual-Leone, A., Michel, C. M., and Farzan, F. (2015). Microstates in resting-state EEG: Current status and future directions. *Neuroscience & Biobehavioral Reviews*, 49:105–113.

Koenig, T., Kottlow, M., Stein, M., and Melie-García, L. (2011). Ragu: a free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. *Computational Intelligence and Neuroscience*, 2011:4.

Koenig, T., Prichep, L., Lehmann, D., Sosa, P. V., Braeker, E., Kleinlogel, H., Isenhart, R., and John, E. R. (2002). Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. *Neuroimage*, 16(1):41–48.

Krzanowski, W. and Lai, Y. (1988). A criterion for determining the number of groups in a dataset using sum of squares clustering. *Biometrics*, 44(1):23–34.

Langer, N., Ho, E. J., Alexander, L. M., Xu, H. Y., Jozanovic, R. K., Henin, S., Petroni, A., Cohen, S., Marcelle, E. T., Parra, L. C., et al. (2017). A resource for assessing information processing in the developing brain using EEG and eye tracking. *Scientific Data*, 4.

Lehmann, D. (1971). Multichannel topography of human alpha EEG fields. *Electroencephalography and clinical neurophysiology*, 31(5):439–449.

Lehmann, D., Faber, P. L., Galderisi, S., Herrmann, W. M., Kinoshita, T., Koukkou, M., Mucci, A., Pascual-Marqui, R. D., Saito, N., Wackermann, J., et al. (2005). EEG microstate duration and syntax in acute, medication-naive, first-episode schizophrenia: a multi-center study. *Psychiatry Research: Neuroimaging*, 138(2):141–156.

Lehmann, D., Ozaki, H., and Pal, I. (1987). EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. *Electroencephalography and clinical neurophysiology*, 67(3):271–288.

Lehmann, D. and Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology*, 48(6):609–621.

Lehmann, D., Strik, W., Henggeler, B., König, T., and Koukkou, M. (1998). Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. Visual imagery and abstract thoughts. *International Journal of Psychophysiology*, 29(1):1–11.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2):129–137.

Michel, C. M. and Koenig, T. (2017). EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *NeuroImage*.

Michel, C. M., Koenig, T., Brandeis, D., and Gianotti, L. R. (2009). *Electrical neuroimaging*. Cambridge University Press.

Milz, P. (2016). Keypy – An open source library for EEG microstate analysis.

Murray, M. M., Brunet, D., and Michel, C. M. (2008). Topographic ERP analyses: A step-by-step tutorial review. *Brain Topography*, 20(4):249–264.

Musso, F., Brinkmeyer, J., Mobascher, A., Warbrick, T., and Winterer, G. (2010). Spontaneous brain activity and EEG microstates. A novel EEG/fMRI analysis approach to explore resting-state networks. *Neuroimage*, 52(4):1149–1161.

Ott, C. G. M., Langer, N., Oechslin, M. S., Meyer, M., and Jäncke, L. (2011). Processing of voiced and unvoiced acoustic stimuli in musicians. *Frontiers in psychology*, 2:195.

Pascual-Marqui, R., Michel, C., and Lehmann, D. (1995). Segmentation of brain electrical activity into microstates: model estimation and validation. *Biomedical Engineering, IEEE Transactions on*, 42(7):658–665.

Pedroni, A., Gianotti, L. R., Koenig, T., Lehmann, D., Faber, P., and Knoch, D. (2017). Temporal characteristics of EEG microstates mediate trial-by-trial risk taking. *Brain topography*, 30(1):149–159.

Rokach, L. and Maimon, O. (2005). Clustering methods. In *Data mining and knowledge discovery handbook*, pages 321–352. Springer.

Schiller, B., Gianotti, L. R., Baumgartner, T., Nash, K., Koenig, T., and Knoch, D. (2016). Clocking the social mind by identifying mental processes in the IAT with electrical neuroimaging. *Proceedings of the National Academy of Sciences*, 113(10):2786–2791.

Skrandies, W. (1989). Data reduction of multichannel fields: global field power and principal component analysis. *Brain Topography*, 2(1-2):73–80.

Tibshirani, R. and Walther, G. (2005). Cluster Validation by Prediction Strength. *Journal of Computational and Graphical Statistics*, 14(December 2013):511–528.

Van de Ville, D., Britz, J., and Michel, C. M. (2010). EEG microstate sequences in healthy humans at rest reveal scale-free dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 107(42):18179–84.

Wackermann, J., Lehmann, D., Michel, C., and Strik, W. (1993). Adaptive segmentation of spontaneous EEG map series into spatially defined microstates. *International Journal of Psychophysiology*, 14(3):269–283.

Yuan, H., Zotev, V., Phillips, R., Drevets, W. C., and Bodurka, J. (2012). Spatiotemporal dynamics of the brain at rest—exploring EEG microstates as electrophysiological signatures of BOLD resting state networks. *Neuroimage*, 60(4):2062–2072.

A Optimised iteration scheme for modified K-means

The modified K-means algorithm is based on the loss function

$$E = \sum_{n=1}^{N} ||\mathbf{x}_n - \sum_{k=1}^{K} \mathbf{a}_k z_{k,n}||^2 \tag{24}$$

which should be minimised under the constraints

$$\begin{aligned} z_{k,n} \cdot z_{k',n} &= 0 \quad \text{if} \quad k \neq k' \\ ||\mathbf{a}_k||^2 &= 1. \end{aligned} \tag{25}$$

The first constraint ensures that a single microstate is active at any given time, while the second constraint resolves the scaling ambiguity between the $\mathbf{a}$ and the $\mathbf{z}$ variables. While closed form optimization is not feasible, we can consider two different iterative schemes.

We can follow a K-means approach as in Pascual-Marqui et al. (1995): Interchange hard assignment and $\mathbf{a}$ updates. In this case assume the assignments (each data point is assigned to a component l_n) to be given. In this case we may first optimise the loss function w.r.t. the $z_{l_n,n}$ variables

$$\widehat{z_{l_n,n}} = \arg\min_z ||\mathbf{x}_n - \mathbf{a}_{l_n} z||^2 = \mathbf{a}_{l_n}^\intercal \mathbf{x}_n \tag{26}$$

Substitute this activation measure in the loss function to obtain

$$E = \sum_{n=1}^{N} ||\mathbf{x}_n - \mathbf{a}_{l_n}(\mathbf{a}_{l_n}^\intercal \mathbf{x}_n)||^2 = \sum_{n=1}^{N} ||\mathbf{x}_n||^2 - (\mathbf{a}_{l_n}^\intercal \mathbf{x}_n)^2 \tag{27}$$

This loss function can now be minimised w.r.t. $\mathbf{a}_k$ under the unit length constraint, leading to the eigenvalue problem

$$\left(\sum_{n=1}^{N} \delta_{l_n,k} \mathbf{x}_n \mathbf{x}_n^\intercal \right) \widehat{\mathbf{a}_k} = \lambda_1 \widehat{\mathbf{a}_k} \tag{28}$$

Where the delta function weighted sum $\sum_{n=1}^{N} \delta_{l_n,k}$ assures that only data points associated with the component, i.e., $l_n = k$ are included.

Alternatively, we may consider an iterative scheme in which we first keep $z_{k,n}$ fixed (i.e., both the assignments l_n and the activities $z_{l_n,n}$) and optimise for the scalp maps $\mathbf{a}$. This leads to

$$\begin{aligned} \mathbf{a}_k &= \arg\min_{\mathbf{a}_k} \sum_{n=1}^{N} \delta_{l_n,k} ||\mathbf{x}_n - \mathbf{a}_k z_{k,n}||^2 \quad \text{s.t.} \quad ||\mathbf{a}_k||^2 = 1 \\ \Rightarrow \quad \mathbf{a}_k &= \frac{\sum_{n=1}^{N} \delta_{l_n,k} \mathbf{x}_n z_{k,n}}{||\sum_{n=1}^{N} \delta_{l_n,k} \mathbf{x}_n z_{k,n}||}. \end{aligned} \tag{29}$$

Next we obtain new values for l_n and $z_{k,n}$ similar to the first approach.

$$\begin{aligned} \widehat{z_{l_n,n}} &= \arg\min_z ||\mathbf{x}_n - \mathbf{a}_{l_n} z||^2 = \mathbf{a}_{l_n}^\intercal \mathbf{x}_n \\ l_n &= \arg\min_k (||\mathbf{x}_n||^2 - (\mathbf{a}_k^\intercal \mathbf{x}_n)^2). \end{aligned} \tag{30}$$

The two approaches lead to different updates for $\mathbf{a}$. The second scheme is slightly faster than the first and turns out to produce an improved fit when empirically evaluated, see Figure 11.

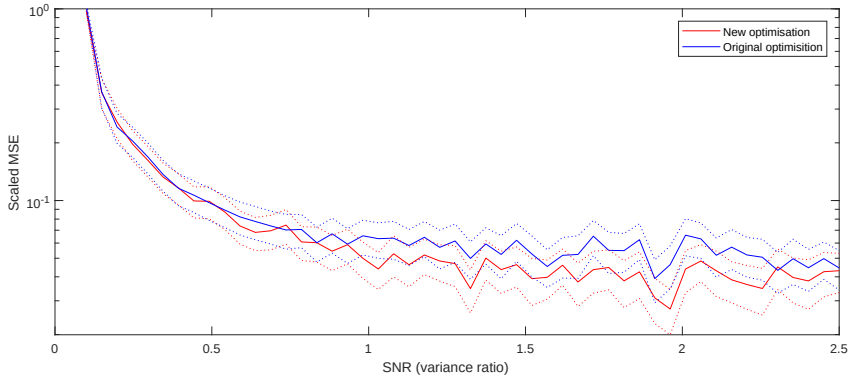


Figure 11: Simulation experiment to benchmark the two optimization procedures for modified K-means. Simulated data ($N = 2000$, $K = 10$, $C = 30$) and variable signal-to-noise rations (variance ratios) were analysed using $N_{\text{its}} = 50$ iterations (typical convergence took about 20 iterations). A scaled mean square error (MSE) measure is computed between the noise free signal, $\mathbf{X}_0$, and the reconstructed signal, $\mathbf{AZ}^\top$, as $\left\|\mathbf{X}_0 - \mathbf{AZ}^\top\right\|^2 / \|\mathbf{X}_0\|^2$. For this simulated data, the new iteration scheme were on average $\sim$4 times faster than the original.

B MATLAB code to Tutorial

```
%% 3 Tutorial: EEG microstates analysis on spontaneous EEG Data
%
% This script executes the analysis steps of the Tutorial described in
% detail in section 3.3 to 3.8 of:
% Poulsen, A. T., Pedroni, A., Langer, N., & Hansen, L. K. (2018).
% Microstate EEGlab toolbox: An introductionary guide. bioRxiv.
%
% Authors:
% Andreas Trier Poulsen, atpo@dtu.dk
% Technical University of Denmark, DTU Compute, Cognitive systems.
%
% Andreas Pedroni, andreas.pedroni@uzh.ch
% University of Zurich, Psychologisches Institut, Methoden der
% Plastizitaetsforschung.

clear;clc;

% start EEGLAB to load all dependent paths
eeglab

%% set the path to the directory with the EEG files
% change this path to the folder where the EEG files are saved
EEGdir = '~/EEGFiles/';

% retrieve a list of all EEG Files in EEGdir
EEGFiles = dir([EEGdir '*.set']);

%% 3.3 Data selection and aggregation

%% 3.3.1 Loading datasets in EEGLAB
for i=1:length(EEGFiles)
    EEG = pop_loadset('filename',EEGFiles(i).name,'filepath',EEGdir);
```

```
    [ALLEEG, EEG, CURRENTSET] = eeg_store( ALLEEG, EEG, 0 );
    eeglab redraw % updates EEGLAB datasets
end

%% 3.3.2 Select data for microstate analysis
[EEG, ALLEEG] = pop_micro_selectdata( EEG, ALLEEG, 'datatype', 'spontaneous',...
    'avgref', 1, ...
    'normalise', 0, ...
    'MinPeakDist', 10, ...
    'Npeaks', 1000, ...
    'GFPthresh', 1, ...
    'dataset_idx', 1:4 );

% store data in a new EEG structure
[ALLEEG EEG] = eeg_store(ALLEEG, EEG, CURRENTSET);
eeglab redraw % updates EEGLAB datasets

%% 3.4 Microstate segmentation
% select the "GFPpeak" dataset and make it the active set
[ALLEEG EEG CURRENTSET] = pop_newset(ALLEEG, EEG, 4,'retrieve',5,'study',0);

eeglab redraw


% Perform the microstate segmentation
EEG = pop_micro_segment( EEG, 'algorithm', 'modkmeans', ...
    'sorting', 'Global explained variance', ...
    'Nmicrostates', 2:8, ...
    'verbose', 1, ...
    'normalise', 0, ...
    'Nrepetitions', 50, ...
    'max_iterations', 1000, ...
    'threshold', 1e-06, ...
    'fitmeas', 'CV',...
    'optimised',1);

[ALLEEG EEG] = eeg_store(ALLEEG, EEG, CURRENTSET);

%% 3.5 Review and select microstate segmentation

%% 3.5.1 Plot microstate prototype topographies
figure;MicroPlotTopo( EEG, 'plot_range', [] );

%% 3.5.2 Select active number of microstates
EEG = pop_micro_selectNmicro( EEG);
[ALLEEG EEG] = eeg_store(ALLEEG, EEG, CURRENTSET);

%% Import microstate prototypes from other dataset to the datasets that should be
    back-fitted
% note that dataset number 5 is the GFPpeaks dataset with the microstate
% prototypes
for i = 1:length(EEGFiles)
    fprintf('Importing prototypes and backfitting for dataset %i\n',i)
    [ALLEEG EEG CURRENTSET] = pop_newset(ALLEEG, EEG, CURRENTSET,'retrieve',i,'study',0);
    EEG = pop_micro_import_proto( EEG, ALLEEG, 5);
```

```
    %% 3.6 Back-fit microstates on EEG
    EEG = pop_micro_fit( EEG, 'polarity', 0 );

    %% 3.7 Temporally smooth microstates labels
    EEG = pop_micro_smooth( EEG, 'label_type', 'backfit', ...
        'smooth_type', 'reject segments', ...
        'minTime', 30, ...
        'polarity', 0 );

    %% 3.9 Calculate microstate statistics
    EEG = pop_micro_stats( EEG, 'label_type', 'backfit', ...
        'polarity', 0 );

    [ALLEEG EEG] = eeg_store(ALLEEG, EEG, CURRENTSET);

end


%% 3.8 Illustrating microstate segmentation
% Plotting GFP of active microstates for the first 1500 ms for subject 1.
[ALLEEG EEG CURRENTSET] = pop_newset(ALLEEG, EEG, CURRENTSET,'retrieve',1,'study',0);
figure;MicroPlotSegments( EEG, 'label_type', 'backfit', ...
    'plotsegnos', 'first', 'plot_time', [4200 5700], 'plottopos', 1 );

eeglab redraw
```

APPENDIX C

Neural Markers of Responsiveness to the Environment in Human Sleep

Authors: Thomas Andrillon, **Andreas Trier Poulsen**, Lars Kai Hansen, Damien Léger, and Sid Kouider.

Status: Published in *Journal of Neuroscience.*

URL: http://www.jneurosci.org/content/36/24/6583

Behavioral/Cognitive

Neural Markers of Responsiveness to the Environment in Human Sleep

Thomas Andrillon,[1,2] Andreas Trier Poulsen,[3] Lars Kai Hansen,[3] Damien Léger,[4] and Sid Kouider[1]

[1]Brain and Consciousness Group, Département d'Études Cognitives, École Normale Supérieure, PSL Research University, 75005 Paris, France, [2]École Doctorale Cerveau Cognition Comportement, Université Pierre et Marie Curie, 75005 Paris, France, [3]Technical University of Denmark, DTU Compute, 2800 Lyngby, Denmark, and [4]Université Paris Descartes, Sorbonne Paris Cité, APHP, Hôtel Dieu, Centre du Sommeil et de la Vigilance et EA 7330 VIFASOM, 75004 Paris, France

Sleep is characterized by a loss of behavioral responsiveness. However, recent research has shown that the sleeping brain is not completely disconnected from its environment. How neural activity constrains the ability to process sensory information while asleep is yet unclear. Here, we instructed human volunteers to classify words with lateralized hand responses while falling asleep. Using an electroencephalographic (EEG) marker of motor preparation, we show how responsiveness is modulated across sleep. These modulations are tracked using classic event-related potential analyses complemented by Lempel-Ziv complexity (LZc), a measure shown to track arousal in sleep and anesthesia. Neural activity related to the semantic content of stimuli was conserved in light non-rapid eye movement (NREM) sleep. However, these processes were suppressed in deep NREM sleep and, importantly, also in REM sleep, despite the recovery of wake-like neural activity in the latter. In NREM sleep, sensory activations were counterbalanced by evoked down states, which, when present, blocked further processing of external information. In addition, responsiveness markers correlated positively with baseline complexity, which could be related to modulation in sleep depth. In REM sleep, however, this relationship was reversed. We therefore propose that, in REM sleep, endogenously generated processes compete with the processing of external input. Sleep can thus be seen as a self-regulated process in which external information can be processed in lighter stages but suppressed in deeper stages. Last, our results suggest drastically different gating mechanisms in NREM and REM sleep.

Key words: complexity; EEG; NREM; REM; sensory processing; sleep

Significance Statement

Previous research has tempered the notion that sleepers are isolated from their environment. Here, we pushed this idea forward and examined, across all sleep stages, the brain's ability to flexibly process sensory information, up to the decision level. We extracted an EEG marker of motor preparation to determine the completion of the sensory processing chain and explored how it is constrained by baseline and evoked neural activity. In NREM sleep, slow waves elicited by stimuli appeared to block response preparation. We also used a novel analytic approach (Lempel-Ziv complexity) and showed that the ability to process external information correlates with neural complexity. A reversal of the correlation between complexity and motor indices in REM sleep suggests drastically different gating mechanisms across sleep stages.

Introduction

Sleep can be defined as a state of behavioral unresponsiveness (Peigneux et al., 2001), but its extent and underlying mechanisms need to be further specified. Disconnection from the external world may play a crucial role in memory consolidation, allowing the brain to turn inward and protect endogenous mechanisms of neural plasticity from external interferences (Rasch and Born, 2013). Yet, this disconnection potentially comes at the expense of survival, as a sleeping organism becomes highly vulnerable. It remains unclear, however, to which extent such disconnection is implemented and modulated. The influential thalamic gating

Received March 18, 2016; revised May 11, 2016; accepted May 14, 2016.

Author contributions: T.A. and S.K. designed research; T.A., D.L., and S.K. performed research; T.A., A.T.P., and L.K.H. analyzed data; T.A., A.T.P., L.K.H., D.L., and S.K. wrote the paper.

This work was supported by Agence Nationale de la Recherche Grants ANR-10-LABX-0087 and ANR-10-IDEX-0001-02, European Research Council Project METAWARE to S.K., and Ministère de la Recherche and the Société Française de Recherche et Médecine du Sommeil to T.A. We thank Hernan Anllo, Leonardo Barbosa, Virginie Bayon, Audrey Dalbin, Isabelle Dautriche, Livio de Sanctis, Maxime Elbaz, Louise Goupil, Stéphane Rio, and Chiara Varazzani for their help.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Thomas Andrillon, Brain and Consciousness Group, Département d'Études Cognitives, 29 rue d'Ulm, École Normale Supérieure, PSL Research University, 75005 Paris, France. E-mail: thomas.andrillon@gmail.com.

DOI:10.1523/JNEUROSCI.0902-16.2016

 0270-6474/16/366583-14$15.00/0

hypothesis proposed that behavioral unresponsiveness was achieved at an early stage, through the blockade of sensory information at the thalamic level (McCormick and Bal, 1994). However, familiar and salient stimuli trigger an awakening more easily (Oswald et al., 1960; Formby, 1967), suggesting the preservation of basic sensory processes. Even when sleep is preserved, familiar stimuli exhibit different brain responses (Perrin et al., 1999). In recent years, several studies have shown that external information processing may extend far beyond automatic operations: from the detection of semantic incongruity (Bastuji and García-Larrea, 1999; Ibáñez et al., 2006) or the violation of simple rules (Ruby et al., 2008; Strauss et al., 2015) to the formation of new associations (Arzi et al., 2012; de Lavilléon et al., 2015). The cognitive processes involved during sleep therefore appear more elaborated than previously thought.

However, although these studies have demonstrated the preservation of cognitive abilities during sleep, the neural mechanisms allowing sleepers to process or isolate from external stimulations remain unsettled. The role of evoked slow oscillations, such as K-complexes, a hallmark of NREM sleep, is particularly unclear (Halász, 2005). K-complexes have sometimes been related to arousal systems (Siclari et al., 2014) providing windows of wakefulness to sleepers (Destexhe et al., 2007). Conversely, they have also been described as local down states (Cash et al., 2009) entailing large-scale neuronal silencing (Vyazovskiy and Harris, 2013) and protecting sleep against external stimulations (Wauquier et al., 1995; Bastien et al., 2000). This tendency neurons have to alternate between periods of activations (up states) and silencing (down states) (Sanchez-Vives and McCormick, 2000) is termed "neuronal bistability" and has been proposed as a mechanism limiting sensory processing at the cortical level (Tononi and Massimini, 2008; Pigorini et al., 2015). Sleep spindles are also thought to enable sleeper's isolation by blocking incoming input at the thalamic level (McCormick and Bal, 1994). However, evidence supporting this hypothesis is scarce (Schabus et al., 2012), whereas cellular recordings showed remarkably preserved responses to sounds during spindles (Sela et al., 2016).

REM sleep, on the other hand, has been far less studied. Paradoxically, even if brain activity in REM sleep resembles wakefulness and consciousness is regained (Hobson and Pace-Schott, 2002), sleepers remain largely unresponsive (Ermis et al., 2010). It has been proposed that dreams themselves would compete and block the processing of external inputs (Nir and Tononi, 2010), but direct evidence is still missing.

Here, to explore responsiveness during human sleep, we relied on a paradigm that aimed at inducing task-dependent responses in the sleeping brain during naps (Kouider et al., 2014). In practice, participants fell asleep while categorizing spoken words with lateralized responses. Using EEG recordings, we explored whether stimuli elicited brain activations corresponding to the motor preparation of the correct response (i.e., lateralized readiness potential [LRP]) (Masaki et al., 2004; Smulders et al., 2012). The presence of such LRP implies that stimuli were not only encoded at the sensory level but also processed at a high level of semantic representation and propagated up to the preparation of a motor response (see Kouider et al., 2014). The presence of an LRP highlights the maintenance of complex distributed processes during sleep. Although our previous study was limited to daytime sleep (nap study), here we investigated all sleep stages in a full-night protocol. Furthermore, the presence and magnitude of the LRP was further evaluated in light of brain activity before and after the stimulus so as to provide novel insights on how the brain manages to alternatively track its environment or rather preserve sleep.

Materials and Methods

Participants

Twenty-three right-handed French native speakers (16 females, age 21–31 years) with no history of neurological or sleep disorders participated to this study. Participants underwent an interview with a sleep specialist and filled in questionnaires determining their sleep habits and their propensity to fall asleep in noisy environments. Participants did not complain of any sleep disorder. They were monitored for 7–10 d before the recording session through actigraphy and sleep diaries to ensure regular sleep/wake rhythms. Among these 23 participants, 5 participants (4 females) were discarded from our analyses either for technical issues affecting the recordings ($N = 1$) or because the experiment was aborted ($N = 4$, participants experiencing difficulties to fall asleep with auditory stimulations). The protocol had been approved by the local ethics committee (Comité de Protection des Personnes, Ile-de-France I, Paris, France).

Experimental procedure

Task. We adapted a procedure previously used during daytime naps (Kouider et al., 2014) to a full-night protocol to explore all sleep stages. On the night of the recordings, participants were equipped for polysomnographic recordings. They went to bed, and spoken words in French were then played in isolation one after the other. These words referred either to an animal or to an object. Participants were instructed to perform a semantic-decision task by indicating the category of each word through right- and left-hand responses. They were asked to categorize words as long as they were awake and to resume responding in case of an awakening. Subjects were reminded to do so whenever they awoke during the night without resuming to respond. However, participants were explicitly authorized to fall asleep while performing the task. Crucially, three different lists of words were played to participants according to their vigilance state, as assessed through an online assessment of sleep stages (see below). A list was played whenever participants were in NREM sleep, another list was played during REM sleep, and a wake list was played otherwise (i.e., mainly during wakefulness). Thus, unpracticed words (i.e., novel words not previously categorized during wakefulness) were played in sleep to ensure that participants had to access the meaning of the word to prepare for the appropriate response, without relying on stimulus-response associations learned while awake.

Stimuli. Stimuli were French spoken words uttered by a female voice. Lists of 72 words were created each containing 36 words referring to animals and 36 words referring to objects. Animal and object words were matched in frequency and number of syllables using the Lexique database (New et al., 2004). The attribution of the different lists to a given vigilance state was counterbalanced across participants. Participants received on average 21.7 ± 1.0 (mean ± SEM across participants), 25.2 ± 0.7, and 6.9 ± 0.3 times the words from the wake, NREM, and REM lists, respectively.

Apparatus. Response-handles were attached to participants' hands. The mapping between semantic categories and response-hands was counterbalanced across participants. Stimuli were played through loud speakers at ~55 dB (depending on participants' preferences) using the Psychtoolbox extension (Brainard, 1997) for MATLAB (The MathWorks). Stimuli were played every 6–9 s (random uniform jitter).

Data analysis

Electrophysiological recordings. EEG ($N = 19$ derivations, 10–20 montage), electro-occulographic (EOG, $N = 2$ derivations, positioned above the right canthus and under the left canthus), electromyograhpic (EMG, 1 derivation placed on the chin measuring muscle tone and 2 derivations on the right and left abductor pollicis brevis [thumb flexor muscle] recording muscle activity accompanying participants' responses) and electrocardiographic (ECG, $N = 1$ derivation) data were continuously recorded. Video monitoring was also available. EEG, EOG, ECG, and EMG data were recorded with AgCl electrodes attached to participants' skin and hair with an adhesive paste (EC2, Natus Neurology). Signals

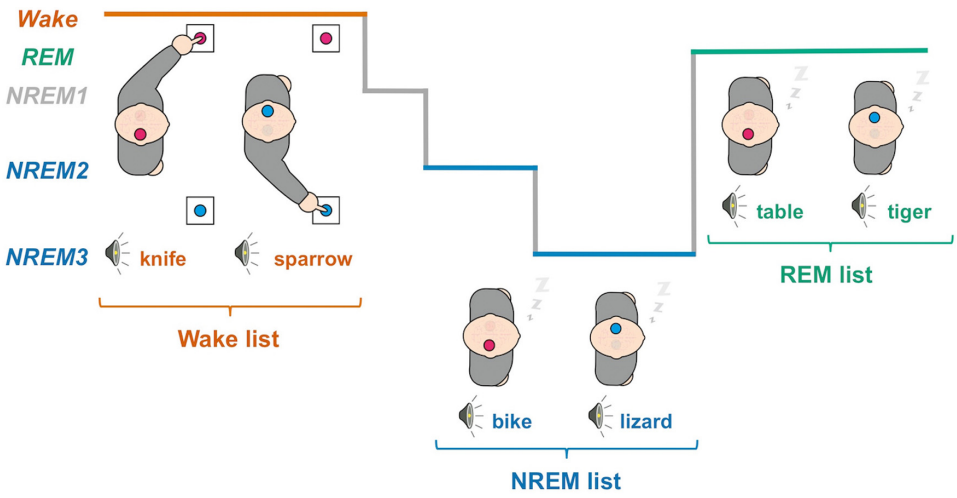


Figure 1. Experimental procedure. Illustration of the protocol. Different lists of animal and object words were played to participants. Participants were instructed to classify these words through left- and right-hand responses according to their semantic category (here right-hand responses for animals). Lateralized hand-response preparation involves the contralateral motor cortices, a task-dependent lateralization of brain activity that can be tracked with the EEG (Fig. 2). Different lists of words were presented to participants. A list was restricted to NREM sleep (NREM2 and NREM3, blue) and another one to REM sleep (green) while the wake list (red) was played otherwise. Changing list between wake and sleep prevents sleepers from using stimulus-response associations learned in wake to classify words in sleep. Words being novel, participants must have had access to the meaning of each word to prepare for the correct response.

were amplified through a B1IP or B2IP MEDATEC amplifier (Medical Data Technology) and recorded at a 200 Hz sampling rate. Impedances of scalp electrodes were generally <5 kΩ. EEG and EOG electrodes were referenced online to the opposite mastoids. An additional channel was used to synchronize EEG data with stimulus presentation.

Online sleep assessment. The presentation of stimuli was continuous, but the words presented to participants were selected from lists that differed depending on participant's vigilance state (Fig. 1). Vigilance and sleep states were scored online following standard guidelines (Iber et al., 2007). In practice, the experimenter waited for the appearance of NREM hallmarks (first spontaneous K-complex or sleep spindle in the absence of arousal) before switching to the NREM list. Similarly, the REM list was presented to participants when they entered the REM stage (absence of slow oscillations or sleep spindles, absence of alpha oscillations, presence of saw-tooth waves, rapid eye movements, increase in theta oscillations, reduced or absent muscle tone). The experimenter typically waited for a few minutes after the transition to the new sleep stage before switching the list, so as to ensure sleep stage's stability. The wake list was presented to participants whenever they showed signs of arousal (body movements, increase in low-amplitude desynchronized rhythms, EMG activation). This online scoring was confirmed and refined offline (see below).

Offline sleep scoring. Wakefulness and sleep stages were scored by 2 scorers blind to experimental conditions and following established guidelines (Iber et al., 2007). Only Fz, C3, C4, and Pz EEG derivations (within the 10–20 montage) were used along ECG and EMG derivations. EEG channels were first rereferenced to the average mastoids. The EEG and EOG signals were high-pass filtered >0.1 Hz and then low-pass filtered <30 Hz (two-pass Butterworth filters at the fifth order). Using the same filter types, EMG data were bandpass filtered between 60 and 80 Hz. A notch-filter (second order Infinite Impulse Response filter) $\sim$50 Hz was used on all channels to reduce line-noise. Vigilance states were continuously scored on 20-s-long windows as wakefulness, NREM sleep Stages 1, 2, and 3 (NREM1, NREM2, and NREM3 stages, respectively), and REM sleep. Importantly, epochs showing signs of arousal (increase in alpha oscillations or oscillations >16 Hz lasting >3 s) or microarousal (<3 s) in association with trial onsets were marked. The corresponding trials were not included in further analyses to avoid potential confounds. Table 1 summarizes sleep scoring across participants.

Slow wave detection. Slow oscillations (sleep slow waves and K-complexes) were detected in NREM sleep using an algorithm that has been presented in details previously (Riedner et al., 2007; Nir et al., 2011). Briefly, slow oscillations were detected for each EEG channel by bandpass filtering the EEG signal between 0.2 and 3 Hz (two-pass Butterworth filter at the third order). The first-order derivative was used to detect local extrema and identify single waves. Only slow oscillations with a peak-to-peak amplitude exceeding 75 μV and a duration of >0.5 s and <2 s were considered as slow waves. These slow oscillations were used to define light and deep NREM trials (see below).

EEG preprocessing. EEG data were analyzed using a combination of SPM (Functional Imaging Laboratory, University College London, London), FieldTrip (Oostenveld et al., 2011), and EEGlab (Delorme and Makeig, 2004) toolboxes running on MATLAB (The MathWorks). We examined the brain activity in response to sounds by computing event-related potentials (ERPs) and LRPs. To do so, the continuous EEG signal referenced to the average mastoids was high-pass filtered >0.1 Hz (two-pass Butterworth filter at the fifth order) and then segmented on large temporal windows around stimulus onsets ($[-14, 14]$ s). The EEG signal was then low-pass filtered <40 Hz and notch-filtered $\sim$50 Hz (same filter types as for the sleep scoring) to reduce line noise. Epochs were resized to focus on the activity around stimulus onsets ($[-2, 7]$ s). Trials with absolute amplitude over central electrodes (Cz, C3, C4) exceeding a given threshold were discarded (150 μV in wake and REM sleep, and 200 μV in NREM sleep because NREM sleep contains high-amplitude physiological slow waves).

Conditions of interest. In NREM sleep, brain activity varies immensely: from NREM1 in which sleep hallmarks (slow oscillations and sleep spindles) are not yet visible to NREM3 where they predominate (Iber et al., 2007). First, to ensure that observed sensory processes are not due to (even brief) awakenings, NREM1 stage was discarded as well as trials in which any sign of arousal could be observed in the EEG signal. Second, NREM2 and NREM3 trials were separated according to the presence of slow waves associated with stimuli onset. We thus divided trials in four different conditions of interest: (1) wake: correct trials during which participants were scored as awake and the wake list was played; (2) light NREM sleep: trials scored as NREM2 or NREM3, during which participants were unresponsive and in which the NREM list was played but no slow-wave were detected (on a $[-2, 3]$ s window time-locked to stimulus onset); (3) deep NREM sleep: trials scored as NREM2 or NREM3, during which participants were unresponsive and in which the NREM list was played and slow waves were detected (on a $[-2, 3]$ s window time-locked

Table 1. Sleep scoring and conditions of interest[a]

	Wake	NREM1	NREM2	NREM3	REM	Total
Duration (min)	88.5 (±9.6)	45.3 (±5.6)	186.3 (±13.4)	79.7 (±7.3)	76.9 (±6.5)	499.6 (±6.6)
% of trials	17.8 (±1.6)	9.7 (±1.0)	39.6 (±2.0)	16.4 (±1.4)	16.5 (±0.8)	100
Marked arousals	0	12.3 (±1.6)	13.4 (±1.8)	2.6 (±0.5)	7.0 (±1.5)	35.8 (±3.8)

[a]Data are mean (SEM); $N = 18$. Participants' vigilance was scored according to established guidelines as wakefulness, NREM1, NREM2, NREM3, and REM stages on 20-s-long windows (see Materials and Methods). Duration indicates the average time participants spend in each stage. Each trial was attributed a sleep stage defined as the most conservative scoring obtained on a [−2, 8] s window around stimulus onset (% of trials). Marked arousals indicates the mean number of trials that were rescored as wakefulness due to an arousal.

to stimulus onset); (4) REM sleep*:* trials scored as REM sleep, during which participants were unresponsive and in which the REM list was played. Thus, light NREM, deep NREM, and REM trials comprised only words that had not been practiced before, when participants were awake and responsive (no wake list word). For each condition, only participants with at least 70 trials were included. On average, there were 174 ± 20 trials in the wake condition (mean ± SEM across $N = 17$ participants), 232 ± 28 in light NREM sleep ($N = 15$), 400 ± 25 in deep NREM sleep ($N = 18$) and 186 ± 9 in REM sleep ($N = 18$). Light NREM trials occurred almost exclusively in NREM2 (93 ± 1.2%, $N = 15$ participants), whereas deep NREM trials correspond to trials occurring either in NREM2 or NREM3 stages (56% and 44%, respectively, ± 3.5% in 18 participants). Thus, light NREM sleep trials corresponded to a state in which slow oscillations could be observed (NREM2) but not in the vicinity of stimulus onset, whereas deep NREM sleep trials corresponded to a state in which slow oscillations were observed in association with stimuli regardless of the presence of continuous trains of slow waves (e.g., in NREM3). Nonetheless, we obtained qualitatively similar results (see Results) when defining light and deep NREM sleep as NREM2 and NREM3 stages, respectively. We also examined trials corresponding to the definition of light, deep NREM, and REM sleep, but during which the wake list was presented to sleepers instead of the NREM or REM lists. Indeed, participants were switched to the wake list whenever the experimenter was unsure of the state of the sleeper (later verified off-line), at transitions between states or when the experimenter had to attend to another participant (two participants participated in the experiment on each recording night). Although these trials containing practiced words were not randomly intermixed with the unpracticed words, they were scored offline as either REM or NREM sleep by two scorers blind to the list used and revealed ERPs and power spectra nearly identical to the ones obtained in trials in which the sleep lists were played (see Fig. 3). Thus, these trials allowed us to investigate how practiced (i.e., overtly categorized) words were processed during sleep, although precautions must be taken when interpreting such results (see Discussion). Participants with at least 30 trials per conditions were included for this analysis.

ERPs. ERPs were computed by averaging the EEG signal across trials for a given experimental condition after baseline correction ([−0.2, 0] s).

LRPs. LRPs allow the monitoring of action selection and preparation (Smulders et al., 2012). LRPs are usually computed with EEG data time-locked to motor responses but can also be computed time-locked to stimuli (Leuthold, 2003; Töllner et al., 2012). In our case, due to the absence of responses during sleep, we computed LRPs on stimulus-locked data. LRPs were computed using ERPs recorded from the right (C4) and left (C3) electrodes placed over motor cortices as follows:

$$LRP = \frac{(C3_{right-hand} - C3_{left-hand}) + (C4_{left-hand} - C4_{right-hand})}{2}$$

ERPs over C3 and C4 electrodes were computed similarly as described above, except for the baseline correction ([−2, 0] s). Using this formula, LRPs are characterized by a negative deflection starting before participants' response. In subsequent analyses, we extracted the LRP magnitude over the temporal windows in which significant negative clusters were observed (see Figs. 2, 3). To examine how this LRP magnitude was dynamically related to other markers of responsiveness, we computed LRPs on windows of 60 consecutive left- and right-response trials slid every trial (see Figs. 4, 6, 7). The LRP-negative potential was extracted as a positive value (LRP magnitude) by multiplying the LRP formula by −1. A similar operation was performed on the N500 and N550 potentials amplitude (see Figs. 6, 7) to extract their magnitude. Other variables of the EEG signal were estimated on the same windows such as the Lempel-Ziv complexity (LZc; see Fig. 4) or ERP components (see Figs. 6, 7). These variables were *z*-score normalized across trials for each participant and then aggregated across participants to examine their correlation.

Time-frequency analyses. To better understand how sleep rhythms impact sensory processing, we computed the time-frequency decomposition of the EEG signal in response to stimuli. To do so, we applied FFT on 1.28-s-long windows (padding ratio of 2) on the preprocessed EEG signal (see above). The resulting power for each frequency and time was expressed as the log-ratio of the power at the corresponding frequency and time over the baseline activity ([−1.5, 0] s) at the same frequency. LRP magnitude and the activity within the slow waves ([1, 6] Hz) and spindle ([11, 16] Hz) bands were extracted on windows of 60 consecutive right-hand and left-hand responses.

LZc. The LZc measures the complexity of a given signal by estimating its compressibility (Ziv and Lempel, 1977). A temporally unpredictable signal will have a low compressibility and therefore a high LZc value. Such an unpredictable signal will be considered as a highly complex signal. Previous studies have shown that LZc accurately tracks the level of consciousness in patients, healthy subjects under anesthesia and during sleep (Casali et al., 2013; Abásolo et al., 2015; Schartner et al., 2015). For example, when the EEG signal is populated by high-amplitude synchronous slow waves as in deep NREM sleep, the signal becomes more predictable and the associated LZc value decreases compared with wakefulness or REM sleep (see Fig. 4*a*). To calculate the LZc on the continuous EEG, we implemented the approach developed by Schartner et al. (2015). Thus, the complexity was computed at the sensor level. To filter the raw EEG, the MATLAB FDAtool was used to create equiripple Finite Impulse Response FIR filters with linear phase to avoid phase distortion in the EEG. The EEG was first low-pass filtered <85 Hz (111th order) and then notch-filtered ~50 Hz to reduce line noise (296th order). A surface Laplacian was then applied to the EEG data (BCILAB plug-in for the EEGLAB toolbox) (Delorme and Makeig, 2004; Kothe and Makeig, 2013), with a neighbor count of four, to reduce the influence of volume conduction. The data were then epoched around stimuli onset ([−2, 6] s), and the LZc was extracted for each trial by integrating the EEG data over all the sensors on 500-ms-long windows slid every 50 ms. Details of the LZc algorithm can be found previously (Schartner et al., 2015). Briefly, for each window, the mean of the EEG signal was subtracted for each sensor and linear trends over the entire epochs were removed. A Hilbert transform was then applied to the signal to extract its envelope. Each channel was then binarized: values above the mean value of the envelope for the corresponding epoch and sensors were coded as 1, and 0 otherwise. The resulting binary matrix was then reshaped sensor-wise into a vector containing the time-points for all channels. The complexity was computed on such vectors and normalized by the complexity calculated on a randomly shuffled version of the same vector. After normalization, LZc takes values between 1 (minimally compressible, i.e., as predictable as the shuffle data) and 0 (fully compressible, i.e., predictable signal). The normalized complexity was thus computed for each trial and averaged across vigilance states (as scored offline) over the prestimulus baseline ([−1.5, 0] s) to compare the overall level of complexity between these different states (see Fig. 4*a*). We also examined how the complexity was affected by stimulation by averaging the LZc on a [−1.5, 6] s windows for each vigilance state separately in an approach similar to when calculating ERPs (see Fig. 4*b*). To better compare the dynamics, LZc was normalized (ratio) by the level of complexity in the prestimulus window

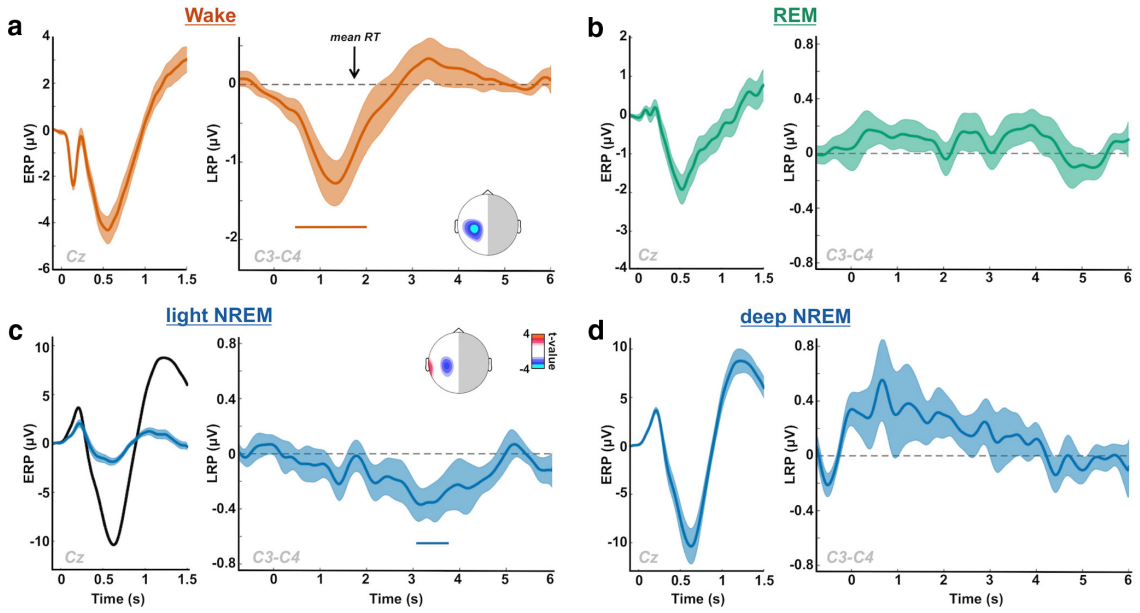


Figure 2. LRPs across sleep stages. The LRP allows monitoring the lateralization of brain activity associated with motor selection and preparation. We used it here as an index of participants' ability to process sensory information up to the semantic level and to use this information in a flexible task-dependent fashion (see Materials and Methods) (Kouider et al., 2014). Subpanels, Left, The stimulus-locked ERP computed on Cz: ***a***, in wakefulness; ***b***, in REM sleep; ***c***, in light NREM sleep (blue, black curves show deep NREM sleep for comparison); ***d***, in deep NREM sleep. Right, The corresponding stimulus-locked LRP computed on C3/C4 electrodes are plotted. Shaded areas represent the SEM computed across participants. Colored horizontal bars represent significant clusters for LRP ($p_{cluster} < 0.05$). Insets, Scalp topographies of LRP averaged over the red and blue clusters. Curves were smoothed using a Gaussian kernel (width: 50 ms for ERPs, 200 ms for LRPs) for display only (statistics were performed before smoothing). An LRP peaking over motor cortices is visible in light NREM but is absent in deeper sleep stages (deep NREM and REM sleep). RT, Response time.

([−1.5, 0] s). To relate the level of complexity with the index of motor preparation (LRP magnitude), the baseline LZc was extracted in wakefulness, NREM and REM sleep on windows of 60 consecutive left- and right hand-response trials slid every trial along the entire recordings. The values obtained were then *z*-scored across trials for each participant before examining their correlation (see Fig. 4*c*).

Sleep cycle identification. To examine how markers of responsiveness are modulated across NREM sleep, we identified the sleep cycles based on each participant's hypnograms. A total of 100 cycles were individualized in 18 participants (5.6 ± 0.2 cycle per participant). Values of interest (LRP magnitude, LZc and δ(<4 Hz) power) were computed on 60 right-hand and left-hand consecutive NREM trials (NREM2 and NREM3) slid every trial within each cycle. Delta power was extracted by applying a FFT of the signal time-locked to stimuli onsets ([−2, 6] s) and normalizing the δ range (<4 Hz) with higher frequencies ([20, 40] Hz, log ratio). Later on, each sleep cycle was normalized in duration to average variables of interest across sleep cycles. To do so, each cycle was divided in 30 equal bins and the mean value of the variables of interests was computed for each bin. Eighty-three cycles were eventually included in this analysis. The other cycles did not have enough NREM2 and NREM3 trials so as to be similarly normalized in duration.

Statistics. To correct for multiple comparisons when examining statistical differences between two time series (see Figs. 2–5), we used a cluster-permutation approach (Maris and Oostenveld, 2007). Each cluster was constituted by the samples that consecutively passed a specific threshold (here, $p < 0.1$ for LRPs and $p < 0.05$ otherwise). As demonstrated previously (Maris and Oostenveld, 2007), this method controls for Type 1 errors independently of this threshold. For each cluster, we computed the sum of the *t* values of all the samples within the cluster. Then, we compared the cluster statistics of each cluster with the maximum cluster statistics of 1000 random permutations and obtained a nonparametric *p* value (Monte-Carlo *p* value: $p_{cluster}$). Significant clusters are displayed as horizontal bars on plots, and $p_{cluster}$ are reported in the main text and figures' legends. For scalp topographies of correlation analyses (Figs. 6, 7), the False Detection Rate (FDR) method was used to correct individual *p* values (Benjamini and Yekutieli, 2011).

Results

Sleepers can classify words during sleep

A classical LRP was observed when participants were awake and responsive (Fig. 2*a*; significant cluster: [0.490, 2.025] s, $p_{cluster} = 0.002$, $N = 17$ participants), characterized by a large negative deflection starting before the average response time (black arrow) over motor electrodes (Fig. 2*a*, inset). When focusing on light NREM sleep, we observed a similar significant negative deflection for the LRP (Fig. 2*c*; significant cluster: [3.035, 3.775] s, $p_{cluster} = 0.01$, $N = 15$ participants), although the corresponding words had not been previously categorized during wakefulness. We obtained a similar LRP even when discarding all words that had been previously presented during an even brief (<3 s) arousal (cluster: [2.830, 3.310] s, $p_{cluster} = 0.043$, $N = 15$ participants), thus when focusing on words that had been presented exclusively during sleep throughout the entire recording session. As in wakefulness, this negative deflection was maximal over motor cortices (central electrodes; see Fig. 2*c*, inset) but was delayed compared with wake trials. This LRP is strikingly similar to the one observed in previous nap studies (Kouider et al., 2014).

However, no significant deviation in the LRP was observed in deep NREM sleep or REM sleep (Fig. 2*b*,*d*), suggesting, at first glance, that processing of sensory information up to the decision level is abolished in deeper stages of sleep. Interestingly, light and deep NREM sleep also differed regarding the evoked responses to stimuli (ERPs). Light NREM sleep was characterized by a reduction of the large negativity evoked by stimuli in NREM sleep (N550) (Picton, 2010). This N550 is often attributed to evoked

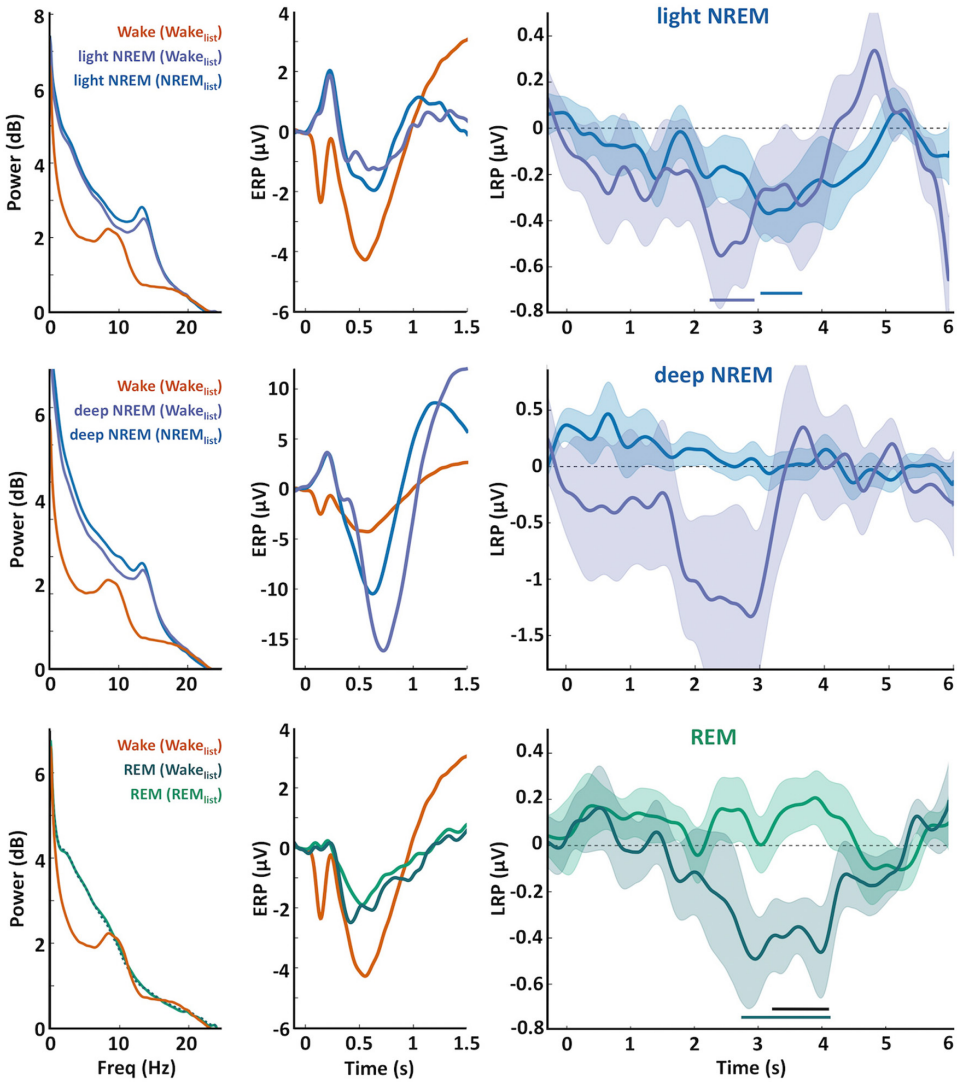


Figure 3. LRPs in light NREM, deep NREM, and REM sleep for words categorized during wakefulness. Top, Power spectra (left), stimulus-locked ERPs (middle), and stimulus-locked LRPs (right) computed in light NREM sleep for trials in which either the NREM list (blue) or the wake (purple) list was played. The wake ERP and power spectrum (red curves) are displayed for comparison. Horizontal bars represent the significant clusters for LRPs ($p_{cluster} < 0.05$). Note the presence of a similar (and slightly earlier) LRP when words categorized during wakefulness were played. Middle, Same plots for deep NREM sleep trials. No significant LRP cluster could be observed for either the NREM or the wake lists. Bottom, Power spectra (left), ERPs (middle), and LRPs (right) computed in REM sleep for trials in which either the REM list (light green) or the wake (dark green) list was played. Interestingly, for words previously categorized in wakefulness (practiced), a clear LRP was observed ($p_{cluster} < 0.05$) but not for unpracticed words (black bar represents cluster for the comparison between the LRPs for the wake and REM lists, $p_{cluster} < 0.05$). Yet, the power spectrum and ERPs for both practiced and unpracticed words are highly similar and different from wake trials (red curves). Shaded areas represent the SEM computed across participants.

slow waves (such as K-complexes) (Bastien et al., 2000), which entails large-scale neuronal silencing (Vyazovskiy and Harris, 2013) and could explain the absence of an LRP when the N550 is larger. Importantly, defining light and deep NREM sleep more classically as NREM2 and NREM3 stages, respectively, led to similar results (i.e., presence of an LRP in light NREM ([4.1, 4.6]s, $p_{cluster} = 0.038$) and absence in deep NREM).

In REM sleep, only words previously heard in wake are classified

Does the absence of an LRP in REM sleep mean that stimuli were not processed at all, or is this absence due to the complexity of the task at hand? To investigate this issue, we took advantage of the fact that words from the wake list were sometimes presented in REM sleep (see Materials and Methods), albeit not intermixed with unpracticed words, which would have allowed a stronger comparison (see Discussion).

Importantly, the ERPs and power spectra were similar when the practiced and unpracticed words were played (Fig. 3), but a clear LRP was observed for the wake list only ([2.750, 3.200] s, $p_{cluster} = 0.036$ and [3.230, 3.900] s, $p_{cluster} = 0.014$, $N = 14$ participants). In addition, a significant difference could be observed between the wake and REM lists ([3.230, 4.120] s, $p_{cluster} = 0.015$). The presence

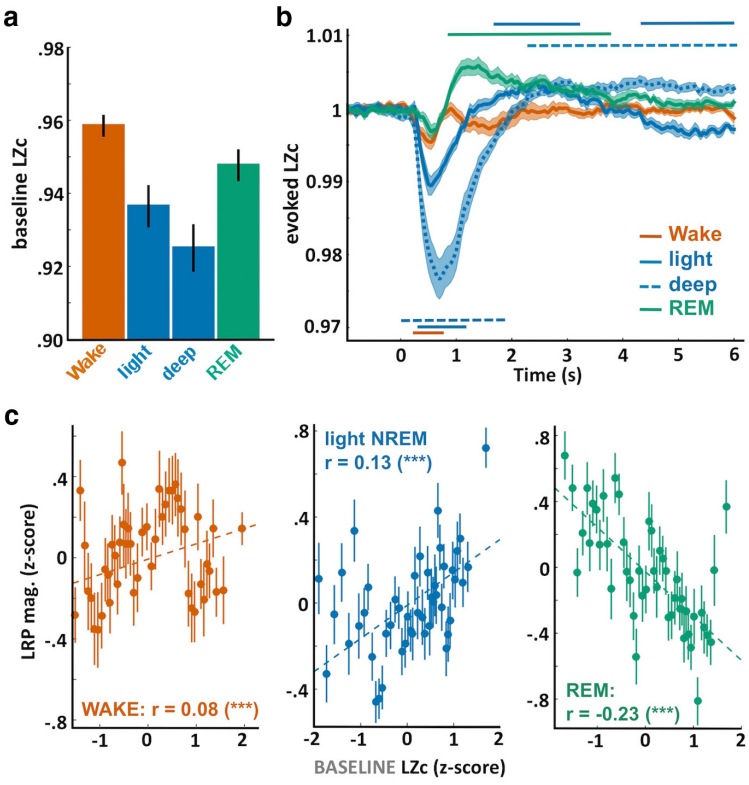


Figure 4. Lempel-Ziv complexity (LZc) across sleep stages and in relation to motor preparation indexes. ***a***, LZc extracted over the prestimulus activity ([−1.5, 0] s) was averaged across trials scored as wakefulness, light NREM, deep NREM, and REM sleep. Error bars indicate the SEM computed across participants. LZc allowed to unambiguously separate the different vigilance states (one-way ANOVA: $F_{(3)} = 9.67, p = 2 \times 10^{-5}, N = 18$ participants). *Post hoc* comparisons show highly significant differences with a gradual decrease in complexity: wake > REM > light NREM > deep NREM (all paired *t* tests, $p < 0.005$). ***b***, LZc time course locked on stimulus onsets and expressed as a ratio of the baseline level ([−1.5, 0] s). Stimuli robustly modulated the complexity of the EEG signal with an initial decrease after stimulus onset ($p_{cluster} < 0.05$, except for REM sleep). The initial decrease was followed by an increase in complexity in light NREM, deep NREM, and REM sleep ($p_{cluster} < 0.05$). ***c***, Correlation between the baseline LZc (see ***a***) and the LRP magnitude computed across the entire night for wake (left), light NREM (middle), and REM sleep (right) trials. Correlation between the pairs of variables was assessed using the Pearson's method, which coefficients are displayed on each subplot along their significance levels. ***$p < 0.005$. Dotted lines indicate the linear fit between the pairs of variables. Values were *z*-scored across trials for each participant before being aggregated across participants. Values were binned for visual purpose ($N = 50$ bins on the sorted LZc values). Error bars indicate the SEM of the LRP magnitude for the corresponding bin.

of an LRP for practiced words shows that sleepers can still map familiar stimuli with the correct response side in REM sleep, suggesting that part of the processing chain was conserved for practiced words. When repeating this procedure in NREM sleep, an LRP was also observed in light NREM for the words practiced in wake ([2.23, 2.9] s, $p_{cluster} = 0.011$, $N = 15$ participants) but not in deep NREM sleep despite the negative deflection visible in Figure 3.

The processing of information depends on the degree of neural complexity

We then set out to examine whether the overall complexity of the prestimulus EEG signal was predictive of sleepers' ability to process information up to the decision level. To do so, we used a recently developed approach (Casali et al., 2013) consisting in measuring the temporal predictability of the EEG signal by reducing it to a single value: the LZc (Ziv and Lempel, 1977). Applying this methodology to our data (Fig. 4*a*) revealed a strong modulation of LZc according to the vigilance state (ANOVA: $F_{(3)} = 9.67, p = 2 \times 10^{-5}$). Such modulation indicates that, in wakefulness, the EEG signal is maximally complex (i.e., unpredictable), whereas, in deep NREM sleep, the EEG signal becomes the most predictable and therefore less complex according to the Lempel-Ziv algorithm. We could separate sleep and wake stages along a gradient matching the phenomenology associated with these states (Nir et al., 2013), confirming initial results in humans (Casali et al., 2013) and animals (Abásolo et al., 2015). Precisely, the baseline LZc was maximal in wakefulness (paired *t* tests across participants with other sleep stages: all $p < 0.005$) and minimal in deep NREM sleep. Light NREM and REM sleep had intermediary values, with REM sleep being the closest to wakefulness (REM vs wake: $t_{(17)} = -3.3$, $p = 0.004$; REM vs light NREM: $t_{(17)} = 5.4, p = 5 \times 10^{-5}$; light vs deep NREM: $t_{(17)} = 9, p = 7 \times 10^{-8}$).

We then investigated whether this baseline LZc was predictive of LRP magnitude. LRP magnitude was extracted on the temporal windows in which a LRP was observed in wakefulness (Fig. 2*a*), light NREM sleep (Fig. 2*c*), and REM sleep (when practiced words were played; Fig. 3). Importantly, the baseline LZc was positively correlated with LRP magnitude in wakefulness and light NREM sleep (Pearsons's coefficient: $r = 0.08$ and 0.13, $p = 0.001$ and 1×10^{-13}, across $N = 2035$ and $N = 3111$ samples in 17 and 15 participants, respectively), but this correlation was reversed in REM sleep ($r = -0.23$, $p = 5 \times 10^{-24}$, $N = 1937$ samples in 18 participants). Comparing LZc in REM sleep when practiced words were presented (LRP present) and when unpracticed words were presented (LRP absent) revealed again that the presence of the LRP was associated with a lower complexity (paired *t* test: $t_{(17)} = 3.5, p = 0.004$). This result contrasts with the correlation observed in NREM sleep and wakefulness, suggesting an inverse relationship between levels of consciousness and the degree of connectedness to the environment (Sanders et al., 2012) in NREM and REM sleep.

We also examined how the LZc was modulated in response to stimuli above and beyond its baseline level (Fig. 4*b*). Stimulus onset was followed by a decrease in LZc (i.e., increase in signal predictability), potentially due to the stereotypical ERPs. And indeed, this decrease in LZc was more pronounced for the sleep stages in which high-amplitude ERPs were observed, such as in NREM sleep (Fig. 2). Accordingly, this decrease in LZc did not reach significance in REM sleep, a state in which auditory stimulation has less impact on brain activity (Bastuji and García-Larrea, 1999; Picton, 2010). Interestingly, in NREM and REM sleep, this initial decrease was followed by an increase in LZc

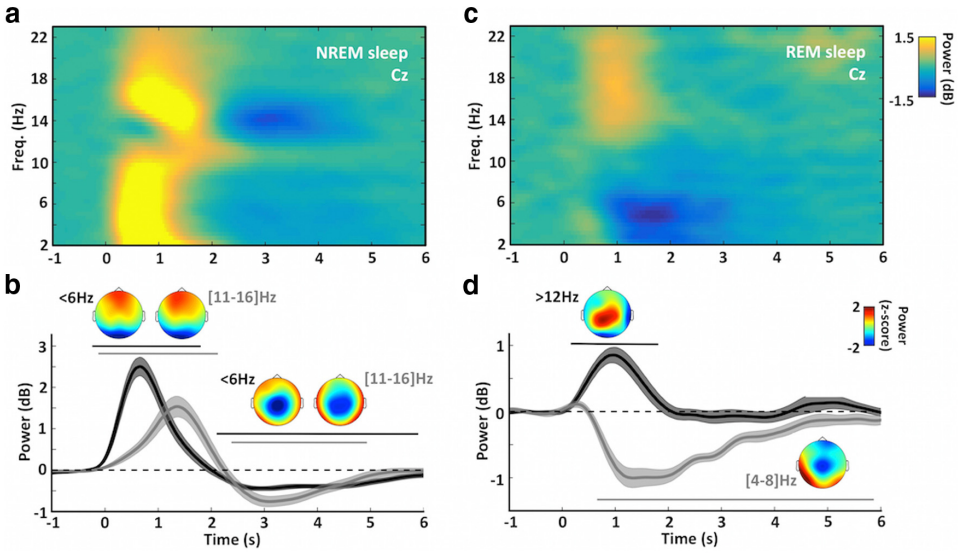


Figure 5. Local modulations of sleep rhythms in association to stimuli. ***a***, Time-frequency decomposition of the EEG signal recorded at Cz in response to stimuli. The time-frequency decomposition was extracted for each trial in NREM sleep (light and deep) and averaged across participants ($N = 18$) (see Materials and Methods). Right after stimulus onset, a large increase in the low-frequency range (<6 Hz) and spindle range ([11, 16] Hz) can be observed, which correspond to slow waves and spindles evoked by stimuli. Interestingly, these sleep rhythms were suppressed later on, at the time during which a LRP was observed in light NREM sleep (Fig. 2). This decrease was confirmed in ***b*** by examining the modulation of the power (at Cz) in these 2 frequency bands (<6 Hz: slow-wave range, black curve; [11, 16] Hz: spindle range, gray curve). Horizontal bars represent the significant clusters determined across participants ($p_{cluster} < 0.05$). Shaded areas represent the SEM computed across participants. Insets, Scalp topographies of the power within the slow-wave and spindle ranges at trial onset ([0, 2] s) and during the LRP window ([2.9, 3.8] s). Power was *z*-scored across sensors to emphasize regional differences. The decrease associated with the LRP is centrally distributed for slow waves and sleep spindles despite their originally frontal distribution, suggesting a local suppression of sleep rhythms. ***c***, ***d***, Same as ***a***, ***b***, except for REM sleep. Note the initial broadband increase in the higher frequency range (>12 Hz) and the decrease within the theta range ([4, 8] Hz). Scalp topographies were computed by averaging power over the significant clusters ($p_{cluster} < 0.05$).

($p_{cluster} < 0.05$), which could be interpreted as a stimulus-induced modulation of sleep depth.

Sensory processes locally disrupt sleep rhythms

So, are the LRPs observed in sleep and the associated increase in LZc reflecting the consequence of a partial awakening? To answer these questions, we examined the consequences of auditory stimulation on sleep itself. Figure 5*a* shows the time-frequency decomposition of the EEG signal in response to sounds in NREM sleep (both light and deep NREM) for electrode Cz. Stimulus onset was accompanied by an increase within the slow-wave (<6 Hz: $p_{cluster} = 0.001$, $N = 18$ participants) and spindle ([11, 16] Hz: $p_{cluster} = 0.002$) ranges (Fig. 5*b*). This increase was followed by a decrease within these two ranges (<6 Hz: [2.1, 6.3] s; [11, 16] Hz: [2.4, 5.0] s, both $p_{cluster} < 0.001$), interestingly, at a time corresponding to the appearance of an LRP in light NREM sleep. This result suggests that the preparation of task-related responses translates into a local-in-time modulation of sleep rhythms and hence of sleep depth. Such modulation was also local-in-space. Indeed, while the increase within the slow-wave and spindle range after stimulus onset showed a typical frontal topography (Fig. 5*b*), the following decrease had a different topography and was prominent over central electrodes (i.e., over the sensors showing the LRP). It is important to note that this decrease within the slow-wave and spindle ranges was not accompanied by an increase in frequency bands associated with wakefulness (alpha: [9, 11] Hz, beta: >16 Hz), which is concordant with the fact that trials with signs of arousal were discarded from our analyses. Thus, in NREM sleep, the processing of auditory information resulted in a local (in time and space) modulation of NREM oscillations in the absence of any observable trace of awakening on the scalp.

In REM sleep, brain responses were qualitatively different (Fig. 5*c*,*d*): auditory inputs clearly disturbed REM sleep with an initial increase in higher frequencies (>12 Hz, $p_{cluster} = 0.01$). This increase was followed by a sustained decrease within the theta range ([4, 8] Hz, $p_{cluster} < 0.001$), reflecting a stimulus-induced perturbation of theta oscillations, which are a hallmark of REM sleep in animals (Buzsáki, 2006). Interestingly, these increases were both maximal over central electrodes where auditory and motor components are observed in wakefulness (Picton, 2010; Smulders et al., 2012). However, these transient markers of arousal did not lead to the appearance of a clear LRP (Fig. 2) unless practiced words were presented (Fig. 3).

Neuronal bistability gates sensory processing in NREM sleep

In NREM sleep, brain responses to stimuli were characterized by strong evoked potentials, such as the P200 and the N550 (Fig. 6*a*). These two potentials are of high interest because the P200 is thought to reflect an activation within the primary sensory cortices corresponding to the eliciting event (i.e., dependent of the sensory modality), whereas the N550 is thought to reflect a broad modality-independent neuronal silencing, maximal in frontal areas (Laurino et al., 2014; Halász, 2015). The respective scalp topographies of these potentials corroborated such view (Fig. 6*a*, right). According to this interpretation, the P200 could represent a marker of the brain's responsiveness to external events, whereas the N550 could be associated with sleep protection against external perturbations.

We thus examined whether the auditory responses (evoked-potentials associated with stimulus onset) were predictive of the appearance of an LRP. Interestingly, in light NREM sleep, the

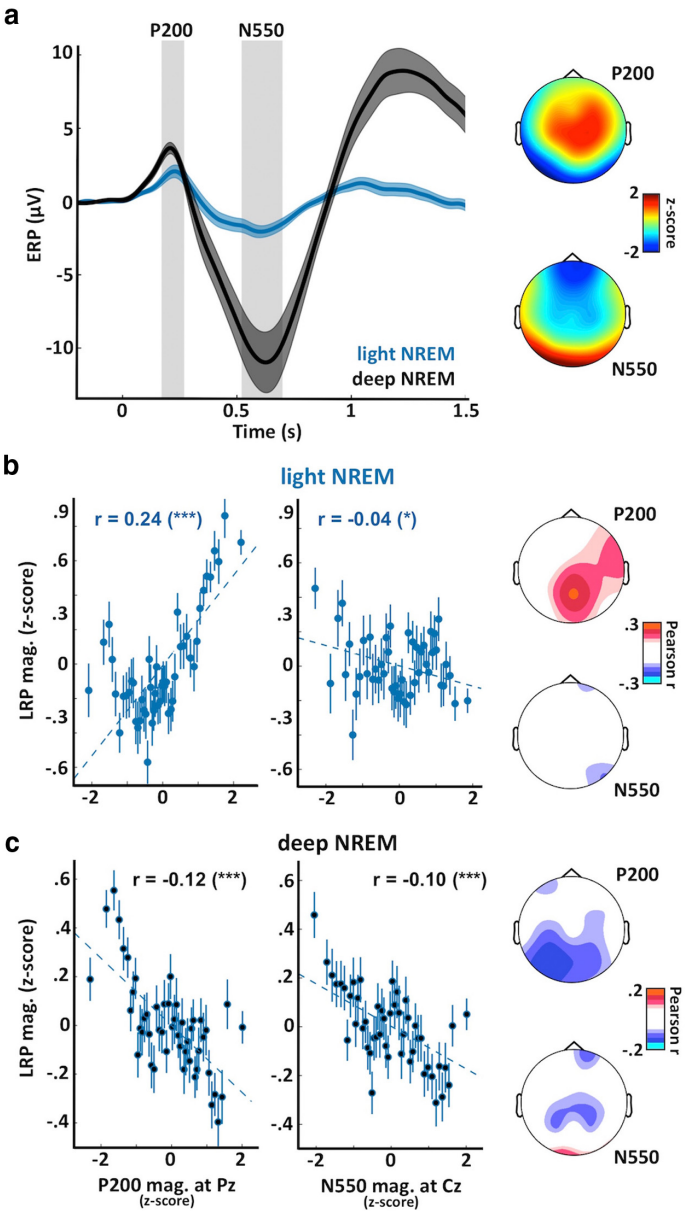


Figure 6. Neural bistability gates sensory processing in NREM sleep. ***a***, ERPs computed at Cz for trials in light NREM (blue curve) and deep NREM sleep (black). Two distinct potentials are clearly visible: a positivity ~200 ms (P200) maximal at centroparietal electrodes (see scalp topography on the top right); and a negativity ~550 ms (N550) predominant in deep NREM sleep (trials associated with slow waves) and maximal at frontal electrodes (see scalp topography on the bottom right). Shaded areas represent the SEM computed across participants ($N = 18$). Scalp topographies were established by averaging the voltage over windows around the two potentials of interest (see gray areas on ERP plot). These values were averaged across participants and *z*-scored across channels to emphasize regional differences. ***b***, Correlations between the LRP magnitude and the P200 magnitude (left) or the N550 magnitude (right, opposite of amplitude) for trials in light NREM sleep. The P200 and N550 magnitudes were computed at Pz and Cz, respectively. Correlation between the pairs of variables was assessed using Pearson's method, with coefficients displayed on each subplot along with their significance levels. ***$p < 0.005$. *$p < 0.05$. Dotted lines indicate the linear fit between the two pairs of variables. Values were *z*-scored across trials for each participant before being aggregated across participants. Values were binned for visual purpose ($N = 50$ bins on the sorted *x*-axis variable). Error bars indicate the SEM of the LRP magnitude for the corresponding bin. Scalp topographies on the right represent the Pearson coefficients computed for each sensor (nonsignificant coefficients were set to 0, $p > 0.05$, FDR corrected for multiple comparisons). ***c***, Same as in ***b*** for trials in deep NREM sleep. The reversal of the relationship between the LRP and the P200 from light to deep NREM paralleled with the appearance of a large N550, showing a suppressive effect on LRP magnitude.

N550 was largely reduced, which is in line with the fact that only trials without slow waves were included. The P200 was also reduced. When examining the relationship between these events and the LRP magnitude in light NREM sleep (Fig. 6*b*), a positive relationship was found between LRP magnitude and the P200 (Pearson coefficient at Pz: $r = 0.24$, $p = 6 \times 10^{-42}$, $N = 3111$ samples in 15 participants). This correlation was maximal over centroparietal electrodes. Thus, the larger the P200, the larger the LRP in light NREM sleep. A much less robust correlation was observed with the magnitude of the N550 ($r = -0.04$, $p = 0.02$), which could be explained by the fact that, in light NREM, the N550 was almost abolished.

On the other hand, in deep NREM sleep, there was a clear negative correlation between the N550 and the LRP magnitude (Fig. 6*c*, Pearson coefficient at Cz: $r = -0.10$, $p = 10^{-4}$, $N = 6037$ samples in 18 participants). Thus, the more pronounced the N550, the smaller the LRP in deep NREM sleep. Interestingly, the relationship between LRP magnitude and the P200 was reversed between light and deep NREM sleep. Contrary to light NREM sleep, the LRP magnitude was negatively correlated with LRP magnitude (Pearson coefficient at Cz: $r = -0.12$, $p = 10^{-20}$) despite the fact that the P200 increases in amplitude from light to deep NREM sleep (Fig. 6*a*), which should favor sensory processing according to the positive relationship observed in light NREM sleep. Such reversal could be due to the fact that the P200 activation can trigger a down state (N550) in deeper stages of NREM sleep, which would ultimately inhibit information processing and reverse the relationship between the P200 and the LRP (see Discussion).

Auditory evoked potentials are predictive of motor preparation in REM sleep

We applied the same approach to REM sleep. Auditory stimuli in REM sleep also evoked archetypal potentials: a positivity ~200 ms followed by a negativity ~500 ms (Fig. 7*a*). However, the topography and temporal profile of these potentials are quite different from NREM potentials. In REM sleep, the P200 is usually associated with the wake auditory P200, generated over the primary and secondary cortices (Picton, 2010). The scalp topography of the P200 with a maximum over central electrodes is concordant with this view. Contrary to light NREM sleep, the P200 was negatively correlated with the

LRP magnitude in REM sleep (Pearson coefficient at Pz: $r = -0.14$, $p = 4 \times 10^{-11}$, $N = 2622$ samples in 18 participants; Fig. 7*b*). This negative correlation was maximal over occipital electrodes.

The N500, a potential that is prominent over occipital electrodes in REM sleep, was positively correlated with LRP magnitude (Pearson coefficient at Pz: $r = 0.15$, $p = 3 \times 10^{-14}$; Fig. 7*c*), an opposite relationship compared with the one found for the N550 potential in NREM sleep. Similar correlations were observed when considering the REM trials in which words practiced in wake were played (P200: $r = -0.12$, $p = 8 \times 10^{-4}$; N500: $r = 0.15$, $p = 7 \times 10^{-15}$). These results indicate that the evoked potentials associated with auditory stimuli are predictive of later and more complex stages of processing (here motor preparation) and suggest that these auditory potentials are not an unspecific reaction to external stimuli but rather reflect cortex's responsiveness to these inputs.

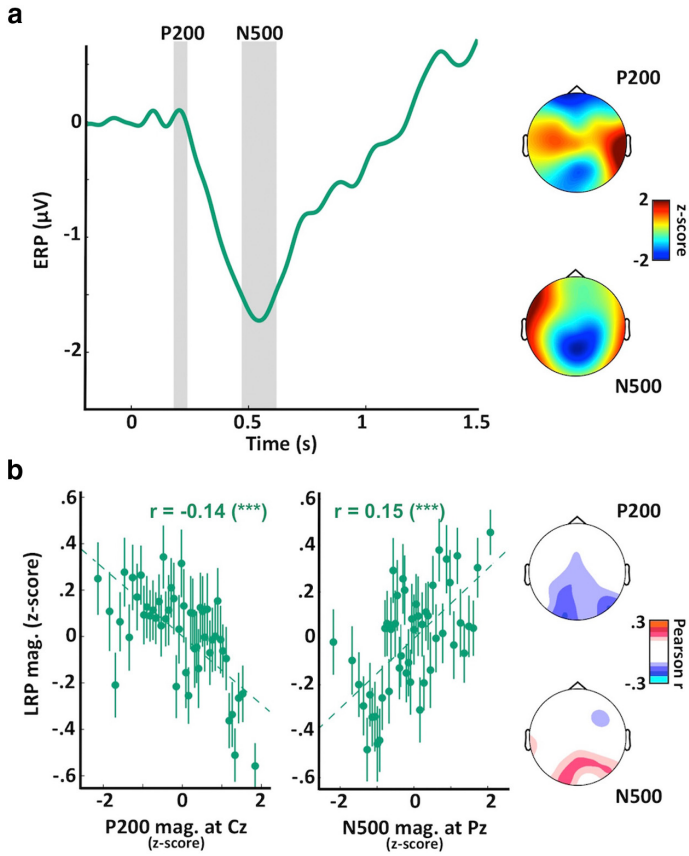


Figure 7. Evoked responses to sounds correlate with LRP magnitude in REM sleep. ***a***, ERPs computed at Cz for trials in REM sleep. Two distinct potentials are again visible: a positivity ~200 ms (P200) maximal at central electrodes (see scalp topography on the top right); and a negativity ~500 ms (N500) maximal at parietal electrodes (see scalp topography on the bottom right). Scalp topographies were established by averaging the voltage over windows around the two potentials of interest (see gray areas on ERP plot). These values were averaged across participants and *z*-scored across channels to emphasize regional differences. These two potentials are quite different from the potentials described in NREM (Fig. 6*a*) in terms of temporal profile, amplitude, and topography. ***b***, Correlations between the LRP magnitude and the P200 magnitude (left) or the N500 magnitude (right, opposite of amplitude) for trials in REM sleep. The P200 and N500 magnitudes were computed at Cz and Pz, respectively. Correlation between the pairs of variables was assessed using Pearson's method, with coefficients displayed on each subplot along with their significance levels. ***$p < 0.005$. Dotted lines indicate the linear fit between the two pairs of variables. Values were *z*-scored across trials for each participant before being aggregated across participants. Values were binned for visual purpose ($N = 50$ bins on the sorted *x*-axis variable). Error bars indicate the SEM of the LRP magnitude for the corresponding bin. Scalp topographies on the right represent the Pearson coefficients computed for each sensor (nonsignificant coefficients were set to 0, $p > 0.05$, FDR corrected for multiple comparisons). Similar correlations were obtained when focusing on practiced words (data not shown).

Markers of responsiveness are modulated within sleep cycles

The presence of LRP in light NREM but not deep NREM sleep or REM sleep (when unpracticed words are presented) could be due to the fact that light NREM sleep is more pervasive to external information. Another interpretation would be that the ability to prepare for the adequate motor response slowly decays with the time spent asleep. With light NREM sleep occurring first, the LRP would be prominent in this state. To test this possibility, we examined how the LRP magnitude was modulated within sleep cycles. Sleep cycles were detected using participant's hypnograms (see Materials and Methods). For each cycle, we retrieved the dynamics of the LZc, the slow-wave power (∂-power, a proxy for slow-wave density), and the LRP magnitude in the NREM part of the cycle (Fig. 8). In accordance with the archetypal profile of sleep cycles, slow-wave power density gradually increased at the beginning of the sleep cycle (descending slope) but decreased toward the end of the cycle (ascending slope and transition to REM sleep) (Halász, 2015). The LZc mirrored this pattern with a gradual decrease in overall complexity followed by a steep increase. Interestingly, the LRP magnitude followed the LZc dynamics and increased again toward the end of the cycle (U shape). This result suggests that the capacity to process information is dynamically related to signal complexity and neural bistability. Thus, the beginning and end of NREM episodes might represent temporal windows in which monitoring of the surrounding environment is possible.

We also examined the effect of the progression within the night on LRP magnitude. To do so, each night was divided into thirds and the LRP magnitude was computed in light NREM, deep NREM, and REM trials in each third separately. In light NREM sleep, a linear decrease was observed across the night (linear regression: $\beta = -0.41$, $p = 0.002$, $N = 15$ participants). A *post hoc* comparison showed a highly significant difference between the LRP magnitude in the first and last third of the night ($t_{(14)} = 4.03$, $p = 0.001$) with only the first third of the night showing a significant LRP ($t_{(14)} = 3.15$, $p = 0.007$). Such within-nights effect was not observed in deep NREM or REM trials (linear regressions: $\beta = -0.20$, $p = 0.44$ and $\beta = -0.08$, $p = 0.58$, respectively). However, because sleep lists were repeated across the night, it is impossible to disentangle in this study a potential effect of early and late sleep from an effect of word-repetition.

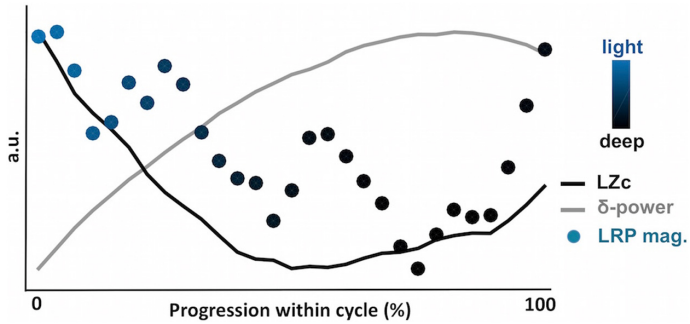


Figure 8. The ability to process information is dynamically modulated within sleep cycles. Modulation of the LRP magnitude (colored dots), the LZc (black curve), and the ∂-power (gray curve) within the NREM sleep part of sleep cycles. Colors of dots (LRP magnitude) represent the proportion of light and deep NREM trials included in the corresponding bin. A classical increase in ∂-power (a proxy for slow-wave density) is observed corresponding to the transition from light to deep NREM sleep. This increase in ∂-power is accompanied by a decrease in LZc and LRP magnitude. Both LZc and LRP have the tendency to increase again at the end of the NREM cycle, paralleling the transition from deep NREM sleep to REM sleep. LRP, LZc, and δ values were estimated within each sleep cycle on fixed windows (see Materials and Methods). Sleep cycles were then binned ($N = 30$) so as to average cycles with different durations, and values were normalized across the entire cycle to better visualize the dynamics of each variable of interest (expressed here in arbitrary units [a.u.]).

Discussion

Complex and distributed processes in the sleeping brain

Participants had to categorize spoken words based on their semantic category through lateralized hand responses while falling asleep. To assess participants' responsiveness, we computed an LRP (Smulders et al., 2012), which indicates whether brain activity is lateralized according to the expected response side. Accordingly, a negative deflection was observed in wakefulness around response time (Fig. 2*a*). This negativity was conserved in light NREM sleep when unpracticed words were played despite the absence of overt responses (Fig. 2*c*), replicating previous findings in naps (Kouider et al., 2014). Because the stimulus-response mapping was counterbalanced across participants, the presence of an LRP for novel words in light NREM sleep reflects the maintenance of complex and distributed processes going from the encoding of the auditory information to the preparation of the appropriate motor response. The maintenance of such complex processing chain could be explained by the automation of the task-set during wakefulness (Kouider et al., 2014).

No LRP was observed in deep NREM sleep or REM sleep (for unpracticed words), suggesting that part of this chain is disrupted in these stages. However, when considering practiced words (wake list, Fig. 1), an LRP was observed in REM sleep (Fig. 3). The presence of an LRP for practiced words only could be interpreted as a failure to access the meaning of novel words during the REM period. Under such circumstances, sleepers could still rely on stimulus-response contingencies learned during wakefulness for practiced words. However, because the practiced and unpracticed words were not intermixed, we cannot rule out the possibility that the presence or absence of an LRP stems from a difference in terms of sleep (i.e., sleep depth) rather than stimulus type (practiced or not).

The complexity of brain dynamics interacts with the processing of the environment

Certain periods of sleep proved more propitious for the induction of task-dependent responses. To understand the relationship between the background neural activity and brain's responsiveness, we computed the temporal predictability of the EEG signal using LZc (Ziv and Lempel, 1977). LZc has been used to determine the level of consciousness during anesthesia or in brain-damaged patients (Casali et al., 2013; Sarasso et al., 2015; Schartner et al., 2015). We confirmed previous findings in sleep (Casali et al., 2013; Abásolo et al., 2015) and LZc matched sleep's phenomenology (Nir et al., 2013), dreams (i.e., conscious contents) being more frequent and complex in REM compared with light NREM but quite rare in deep NREM sleep.

LZc also predicted the LRP magnitude (Fig. 4*c*). In light NREM sleep, an increased propensity to respond to external input was associated with higher levels of complexity before stimulus onset. In REM sleep, however, this relationship was reversed, which could be related to the peculiarity of REM sleep: a state of consciousness disconnected from the environment (Hobson, 2009). This result emphasizes the difference between consciousness and responsiveness as observed in anesthesia (Sanders et al., 2012).

Neuronal bistability and the gating of sensory processing

In contrast with light sleep, no LRP was observed in deep NREM sleep. Light and deep NREM sleep were defined based on the presence or absence of slow oscillations associated with stimuli. Accordingly, light NREM showed a drastically reduced N550 potential (Figs. 2, 6), a potential usually linked to the down states of stimuli-evoked K-complexes (Bastien et al., 2002). Down states represent episodes of neuronal silencing (Steriade, 2003) and perturb the encoding and integration of information (Schabus et al., 2012; Pigorini et al., 2015). The presence of a N550 could therefore prevent the further processing of information and abolish the LRP in deep NREM sleep. This hypothesis is buttressed by the positive correlation between the N550 and LRP magnitudes (Fig. 6*c*).

In light NREM sleep, stimulus-evoked potentials showed an initial P200 correlating positively with LRP magnitude (Fig. 6*b*). The P200 has been interpreted as a local modality-dependent excitation (Laurino et al., 2014) and could reflect brain's responsiveness to external inputs. However, in deep NREM sleep, the P200 was associated with LRP suppression (Fig. 6*c*). We interpret this reversal as the consequence of an increased neuronal bistability from light to deep NREM sleep.

Indeed, in deep NREM sleep, cortical neurons exhibit more down states (Steriade, 2003). Because a local excitation can trigger a down state (Sanchez-Vives et al., 2010; Menicucci et al., 2013), in deep NREM sleep, the P200 could ignite a down state in frontal areas that would shut down in turn any further processing. Thus, depending on sleep depth and neuronal bistability, the relationship of the P200 with sensory processing would reverse: from a potentiating activation to a suppressive mechanism. Brain responses to sounds could therefore follow a self-regulated process (Halász, 2015), whereby sensory activations can be retroactively suppressed. Thus, in NREM sleep, a reactive cortical gating (Tononi and Massimini, 2008), rather than a default thalamic one (McCormick and Bal, 1994), seems to take place.

Could dreams gate sensory processing in REM sleep?

The mechanisms underlying sensory decoupling in REM sleep are still unknown. The recovery of a wake-like brain activity and consciousness should favor the processing of external information, and yet external inputs are rarely integrated in dream scenery (Nir and Tononi, 2010). A recent study described the presence of slow waves in REM sleep restricted to the superficial layers of primary sensory cortices (Funk et al., 2016). These slow waves could act as a gating mechanism because superficial cortical layers are the main targets of thalamic sensory relays (Jones, 2007). We showed here, however, that stimuli had a clear impact on cortical activity: signs of arousals could be observed after stimulus onset (Fig. 5*c*,*d*) as well as robust auditory potentials (Fig. 7). These activations indicate that sensory information did reach the cortex as previously suggested (Sallinen et al., 1996; Nir et al., 2015). However, no LRP was observed when novel words were played. A potential explanation is that stimulus-evoked perturbations were confined in sensory areas.

But what could limit distributed cortical processes if corticocortical connectivity is regained in REM compared to NREM sleep (Massimini et al., 2010)? We propose that sensory information may compete with endogenous contents (i.e., dreams). Indeed, wake-like endogenous activations have been observed in REM sleep (Louie and Wilson, 2001; Andrillon et al., 2015). These endogenous activations could compete for the brain's computational resources and interfere with the processing of external inputs. Accordingly, prestimulus complexity had a negative impact on LRP magnitude (Fig. 4). This "informational gating" could be mediated by a domination of top-down signaling at the expense of bottom-up processes (Nir and Tononi, 2010). Further investigations are needed to investigate the relationship between oneiric contents and sensory disconnection.

Consequences of external stimulations on sleep

How did stimulation affect sleep? In NREM sleep, we observed a two-step response (Fig. 5*a*,*b*). Right after stimulus onset, power increased within the slow-wave and spindle ranges in accordance with the fact that sleep rhythms can be evoked by external stimulations (Halász, 2015). These increases were maximal over frontal electrodes and could correspond to a protective mechanisms ensuring that sleep is preserved (Halász et al., 2014). Likewise, the suppressive N550 (Fig. 6) was maximal at frontal electrodes. After this initial increase, slow waves and spindle power were on the contrary suppressed. The associated spatial and temporal distributions overlap with the time and location of the LRP (Fig. 2*c*) and the increase in LZc following stimulus presentation (Fig. 4*b*).

A local dampening of sleep depth, accompanying sensory processing, thus followed the initial protective response. No sign of arousal was observed, but the absence of arousal at the scalp level does not preclude local awakening in motor or sensory areas (Nobili et al., 2011; Peter-Derex et al., 2015). The local modulation of sleep depth associated with sensory processing could therefore reflect, at the scalp level, the occurrence of local wake within sleep (Nobili et al., 2012). In REM sleep, auditory stimulations also perturbed theta oscillations (Fig. 5*c*,*d*), a hallmark of REM sleep (Buzsáki, 2006). Signs of cortical activation (increase in the high-frequency range) could also be observed after stimulus onset. Once again, these perturbations were maximal over central electrodes, potentially revealing the recruitment of auditory areas.

Responsiveness is dynamically modulated during sleep cycles

We also explored the dynamics of neural complexity, motor preparation, and ∂-power (a proxy for slow-wave density) across sleep cycles. Both the LZc and ∂-power dynamics reflected sleep cycles' structure, with an initial descent toward deep NREM sleep (increase in ∂-power, decrease of LZc) followed by a steep dampening of NREM sleep at the transition with REM sleep. The LRP magnitude was also robustly modulated within sleep cycles and paralleled complexity's dynamics. Crucially, LRP magnitude increased toward the end of sleep cycles. Thus, the LRP does not slowly decay after cycle onset but is tightly linked to sleep depth. This fragility of sleep at the transition with REM sleep could be instrumental in allowing the brain to "decide" whether to stay asleep or to wake up (Halász et al., 2004). This modulation of responsiveness during NREM sleep cycles is reminiscent of a model (Halász and Bódizs, 2013) in which the descending slope of the NREM cycle is gradually dominated by sleep-promoting mechanisms (such as evoked slow waves, as suggested by Fig. 6), whereas the ascending slope is characterized by the emergence of wake-promoting mechanisms, which could explain the recovery of the responsiveness marker toward the end of the NREM episode (Fig. 8). Interestingly, such a model would predict that external information would be better processed at the end of the NREM sleep cycle compared with the beginning, sleep depth being kept equal. Further research would be needed to confirm such a hypothesis.

Overall, our data suggest that sleepers can complexly and flexibly process information during certain sleep stages. This ability is self-regulated by the neural dynamics, allowing the sleeping brain to manage the challenging tradeoff between sensory decoupling and monitoring its environment.

References

Abásolo D, Simons S, Morgado da Silva R, Tononi G, Vyazovskiy VV (2015) Lempel-Ziv complexity of cortical activity during sleep and waking in rats. J Neurophysiol 113:2742–2752. CrossRef Medline

Andrillon T, Nir Y, Cirelli C, Tononi G, Fried I (2015) Single-neuron activity and eye movements during human REM sleep and awake vision. Nat Commun 6:7884. CrossRef Medline

Arzi A, Shedlesky L, Ben-Shaul M, Nasser K, Oksenberg A, Hairston IS, Sobel N (2012) Humans can learn new information during sleep. Nat Neurosci 15:1460–1465. CrossRef Medline

Bastien CH, Ladouceur C, Campbell KB (2000) EEG characteristics prior to and following the evoked K-xomplex. Can J Exp Psychol 54:255–265. Medline

Bastien CH, Crowley KE, Colrain IM (2002) Evoked potential components unique to non-REM sleep: relationship to evoked K-complexes and vertex sharp waves. Int J Psychophysiol 46:257–274. CrossRef Medline

Bastuji H, García-Larrea L (1999) Evoked potentials as a tool for the investigation of human sleep. Sleep Med Rev 3:23–45. CrossRef Medline

Benjamini Y, Yekutieli D (2011) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436. CrossRef Medline

Buzsáki G (2006) Rhythms of the brain. Oxford: Oxford UP.

Casali AG, Gosseries O, Rosanova M, Boly M, Sarasso S, Casali KR, Casarotto S, Bruno MA, Laureys S, Tononi G, Massimini M (2013) A theoretically based index of consciousness independent of sensory processing and behavior. Sci Transl Med 5:198ra105. CrossRef Medline

Cash SS, Halgren E, Dehghani N, Rossetti AO, Thesen T, Wang C, Devinsky O, Kuzniecky R, Doyle W, Madsen JR, Bromfield E, Eross L, Halász P, Karmos G, Csercsa R, Wittner L, Ulbert I (2009) The human K-complex represents an isolated cortical down state. Science 324:1084–1087. CrossRef Medline

de Lavilléon G, Lacroix MM, Rondi-Reig L, Benchenane K (2015) Explicit memory creation during sleep demonstrates a causal role of place cells in navigation. Nat Neurosci 18:493–495. CrossRef Medline

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis

of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21. CrossRef Medline

Destexhe A, Hughes SW, Rudolph M, Crunelli V (2007) Are corticothalamic "up" states fragments of wakefulness? Trends Neurosci 30:334–342. CrossRef Medline

Ermis U, Krakow K, Voss U (2010) Arousal thresholds during human tonic and phasic REM sleep. J Sleep Res 19:400–406. CrossRef Medline

Formby D (1967) Maternal recognition of infant's cry. Dev Med Child Neurol 9:293–298. CrossRef Medline

Funk CM, Honjoh S, Rodriguez AV, Cirelli C, Tononi G (2016) Local slow waves in superficial layers of primary cortical areas during REM sleep. Curr Biol 26:396–403. CrossRef Medline

Halász P (2005) K-complex, a reactive EEG graphoelement of NREM sleep: an old chap in a new garment. Sleep Med Rev 9:391–412. CrossRef Medline

Halász P (2015) The K-complex as a special reactive sleep slow wave: a theoretical update. Sleep Med Rev 29:34–40. CrossRef Medline

Halász P, Bódizs R (2013) Dynamic NREM sleep regulation models. In: Dynamic structure of NREM sleep, pp 7–11. London: Springer.

Halász P, Terzano M, Parrino L, Bódizs R (2004) The nature of arousal in sleep. J Sleep Res 13:1–23. CrossRef Medline

Halász P, Bódizs R, Parrino L, Terzano M (2014) Two features of sleep slow waves: homeostatic and reactive aspects—from long term to instant sleep homeostasis. Sleep Med 15:1184–1195. CrossRef Medline

Hobson JA (2009) REM sleep and dreaming: towards a theory of protoconsciousness. Nat Rev Neurosci 10:803–813. CrossRef Medline

Hobson JA, Pace-Schott EF (2002) The cognitive neuroscience of sleep: neuronal systems, consciousness and learning. Nat Rev Neurosci 3:679–693. CrossRef Medline

Ibáñez A, López V, Cornejo C (2006) ERPs and contextual semantic discrimination: degrees of congruence in wakefulness and sleep. Brain Lang 98:264–275. CrossRef Medline

Iber C, Ancoli-Israel S, Chesson A, Quan S (2007) The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Westchester, IL: American Academy of Sleep Medicine.

Jones EG (2007) The thalamus, Ed 2. Cambridge: Cambridge UP.

Kothe CA, Makeig S (2013) BCILAB: a platform for brain-computer interface development. J Neural Eng 10:056014. CrossRef Medline

Kouider S, Andrillon T, Barbosa LS, Goupil L, Bekinschtein TA (2014) Inducing task-relevant responses to speech in the sleeping brain. Curr Biol 24:2208–2214. CrossRef Medline

Laurino M, Menicucci D, Piarulli A, Mastorci F, Bedini R, Allegrini P, Gemignani A (2014) Disentangling different functional roles of evoked K-complex components: mapping the sleeping brain while quenching sensory processing. Neuroimage 86:433–445. CrossRef Medline

Leuthold H (2003) Programming of expected and unexpected movements: effects on the onset of the lateralized readiness potential. Acta Psychol (Amst) 114:83–100. CrossRef Medline

Louie K, Wilson MA (2001) Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. Neuron 29:145–156. CrossRef Medline

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190. CrossRef Medline

Masaki H, Wild-Wall N, Sangals J, Sommer W (2004) The functional locus of the lateralized readiness potential. Psychophysiology 41:220–230. CrossRef Medline

Massimini M, Ferrarelli F, Murphy M, Huber R, Riedner B, Casarotto S, Tononi G (2010) Cortical reactivity and effective connectivity during REM sleep in humans. Cogn Neurosci 1:176–183. CrossRef Medline

McCormick DA, Bal T (1994) Sensory gating mechanisms of the thalamus. Curr Opin Neurobiol 4:550–556. CrossRef Medline

Menicucci D, Piarulli A, Allegrini P, Laurino M, Mastorci F, Sebastiani L, Bedini R, Gemignani A (2013) Fragments of wake-like activity frame down states of sleep slow oscillations in humans: new vistas for studying homeostatic processes during sleep. Int J Psychophysiol 89:151–157. CrossRef Medline

New B, Pallier C, Brysbaert M, Ferrand L (2004) Lexique 2: a new French lexical database. Behav Res Methods Instrum Comput 36:516–524. CrossRef Medline

Nir Y, Tononi G (2010) Dreaming and the brain: from phenomenology to neurophysiology. Trends Cogn Sci 14:88–100. CrossRef Medline

Nir Y, Staba RJ, Andrillon T, Vyazovskiy VV, Cirelli C, Fried I, Tononi G (2011) Regional slow waves and spindles in human sleep. Neuron 70:153–169. CrossRef Medline

Nir Y, Massimini M, Boly M, Tononi G (2013a) Sleep and consciousness. In: Neuroimaging of consciousness (Cavanna AE, Nani A, Blumenfeld H, Laureys S, eds), pp 133–182. Berlin: Springer.

Nir Y, Vyazovskiy VV, Cirelli C, Banks MI, Tononi G (2015) Auditory responses and stimulus-specific adaptation in rat auditory cortex are preserved across NREM and REM sleep. Cereb Cortex 25:1362–1378. CrossRef Medline

Nobili L, Ferrara M, Moroni F, De Gennaro L, Russo GL, Campus C, Cardinale F, De Carli F (2011) Dissociated wake-like and sleep-like electrocortical activity during sleep. Neuroimage 58:612–619. CrossRef Medline

Nobili L, De Gennaro L, Proserpio P, Moroni F, Sarasso S, Pigorini A, De Carli F, Ferrara M (2012) Local aspects of sleep. In: Progress in brain research, pp 219–232. Amsterdam: Elsevier.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869. CrossRef Medline

Oswald I, Taylor AM, Treisman M (1960) Discriminative responses to stimulation during human sleep. Brain 83:440–453. CrossRef Medline

Peigneux P, Laureys S, Delbeuck X, Maquet P (2001) Sleeping brain, learning brain: the role of sleep for memory systems. Neuroreport 12:A111–A124. CrossRef Medline

Perrin F, García-Larrea L, Mauguière F, Bastuji H (1999) A differential brain response to the subject's own name persists during sleep. Clin Neurophysiol 110:2153–2164. CrossRef Medline

Peter-Derex L, Magnin M, Bastuji H (2015) Heterogeneity of arousals in human sleep: a stereo-electroencephalographic study. Neuroimage 123:229–244. CrossRef Medline

Picton TW (2010) Human auditory evoked potentials. San Diego: Plural.

Pigorini A, Sarasso S, Proserpio P, Szymanski C, Arnulfo G, Casarotto S, Fecchio M, Rosanova M, Mariotti M, Lo Russo G, Palva JM, Nobili L, Massimini M (2015) Bistability breaks-off deterministic responses to intracortical stimulation during non-REM sleep. Neuroimage 112:105–113. CrossRef Medline

Rasch B, Born J (2013) About sleep's role in memory. Physiol Rev 93:681–766. CrossRef Medline

Riedner BA, Vyazovskiy VV, Huber R, Massimini M, Esser S, Murphy M, Tononi G (2007) Sleep homeostasis and cortical synchronization: III. A high-density EEG study of sleep slow waves in humans. Sleep 30:1643–1657. Medline

Ruby P, Caclin A, Boulet S, Delpuech C, Morlet D (2008) Odd sound processing in the sleeping brain. J Cogn Neurosci 20:296–311. CrossRef Medline

Sallinen M, Kaartinen J, Lyytinen H (1996) Processing of auditory stimuli during tonic and phasic periods of REM sleep as revealed by event-related brain potentials. J Sleep Res 5:220–228. CrossRef Medline

Sanchez-Vives MV, McCormick DA (2000) Cellular and network mechanisms of rhythmic recurrent activity in neocortex. Nat Neurosci 3:1027–1034. CrossRef Medline

Sanchez-Vives MV, Mattia M, Compte A, Perez-Zabalza M, Winograd M, Descalzo VF, Reig R (2010) Inhibitory modulation of cortical up states. J Neurophysiol 104:1314–1324. CrossRef Medline

Sanders RD, Tononi G, Laureys S, Sleigh JW (2012) Unresponsiveness ≠ unconsciousness. Anesthesiology 116:946–959. CrossRef Medline

Sarasso S, Boly M, Napolitani M, Gosseries O, Charland-Verville V, Casarotto S, Rosanova M, Casali AG, Brichant JF, Boveroux P, Rex S, Tononi G, Laureys S, Massimini M (2015) Consciousness and complexity during unresponsiveness induced by propofol, xenon, and ketamine. Curr Biol 25:3099–3105. CrossRef Medline

Schabus M, Dang-Vu TT, Heib DP, Boly M, Desseilles M, Vandewalle G, Schmidt C, Albouy G, Darsaud A, Gais S, Degueldre C, Balteau E, Phillips C, Luxen A, Maquet P (2012) The fate of incoming stimuli during NREM sleep is determined by spindles and the phase of the slow oscillation. Front Neurol 3:40. CrossRef Medline

Schartner M, Seth A, Noirhomme Q, Boly M, Bruno MA, Laureys S, Barrett A (2015) Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia. PLoS One 10:e0133532. CrossRef Medline

Sela Y, Vyazovskiy VV, Cirelli C, Tononi G, Nir Y (2016) Responses in rat core auditory cortex are preserved during sleep spindle oscillations. Sleep 39:1069–1082. CrossRef Medline

Siclari F, Bernardi G, Riedner BA, LaRocque JJ, Benca RM, Tononi G (2014) Two distinct synchronization processes in the transition to sleep: a high-density electroencephalographic study. Sleep 37:1621–1637. CrossRef Medline

Smulders FT, Miller JO, Luck SJ, Kappenman E (2012) The lateralized readiness potential. In: Oxford handbook of event-related potential components, pp 209–229. Oxford, UK: Oxford Library of Psychology.

Steriade M (2003) Neuronal substrates of sleep and epilepsy, Ed 1. Cambridge: Cambridge UP.

Strauss M, Sitt JD, King JR, Elbaz M, Azizi L, Buiatti M, Naccache L, van Wassenhove V, Dehaene S (2015) Disruption of hierarchical predictive coding during sleep. Proc Natl Acad Sci U S A 112:E1353–E1362. CrossRef Medline

Töllner T, Rangelov D, Müller HJ (2012) How the speed of motor-response decisions, but not focal-attentional selection, differs as a function of task set and target prevalence. Proc Natl Acad Sci U S A 109:E1990–E1999. CrossRef Medline

Tononi G, Massimini M (2008) Why does consciousness fade in early sleep? Ann N Y Acad Sci 1129:330–334. CrossRef Medline

Vyazovskiy VV, Harris KD (2013) Sleep and the single neuron: the role of global slow oscillations in individual cell rest. Nat Rev Neurosci 14: 443–451. CrossRef Medline

Wauquier A, Aloe L, Declerck A (1995) K-complexes: are they signs of arousal or sleep protective? J Sleep Res 4:138–143. CrossRef Medline

Ziv J, Lempel A (1977) A universal algorithm for sequential data compression. IEEE Trans Inf Theory 23:337–343. CrossRef

APPENDIX D

EEG complexity tracks user failure and difficulty in helicopter simulator

Authors: **Andreas Trier Poulsen**, Jean-maurice Leonetti, Lars Kai Hansen, and Sid Kouider.

Status: Unpublished draft manuscript.

EEG complexity tracks user performance and difficulty in helicopter simulator

Andreas Trier Poulsen[1,*], Jean-maurice Leonetti[2], Lars Kai Hansen[1], and Sid Kouider[2]

[1]Technical University of Denmark, DTU Compute, Kgs. Lyngby, Denmark
[2]Brain and Consciousness Group, Département d'Études Cognitives, École Normale Supérieure, PSL Research University, Paris, France
[*]Corresponding author: atpo@dtu.dk

April 10, 2018

Abstract

Recent research has shown that the complexity of EEG signals is correlated with the levels of cognitive processing in comatose in comatose or sleeping subjects. Here we investigate whether this neural marker can be used in fully aware healthy people, as an index of how focused they are on a given task. We had twenty subjects test a helicopter simulator made for this experiment, with the aim of manoeuvring the helicopter through rings using a joystick. By contrasting when subjects were successful in navigating the helicopter compared to when they made errors we could identify moments of high mental workload and investigate how well EEG complexity was able to track it. Though not able to distinguish on a single-trial level, subjects showed higher complexity on average in the seconds before missing circles compared to when they successfully navigated the helicopter through them. Additionally, a significant drop in complexity was measured in the navigational mode, which subjects reported as the easiest. This mode was presumably the one they improved the most in, reaching a plateau in their improvement early on resulting a decreasing the need to focus, reflected in the decreased complexity. Finally, there was a significant contrast in the complexity for trials with a hard difficulty compared to trials that were easier to navigate. This was also true when only comparing difficulty for successfully completed trials. This indicates that the EEG complexity is not only is sensitive to neural changes leading up to a failure, but also when a subject is struggling during a hard but successful trial. Our results indicate that EEG complexity is able to distinguish moments of varying workload consistently across subjects, which suggests that EEG complexity is a viable candidate as an index of mental workload.

1 Methods

Twenty subjects (hereof 10 were female) were recruited for an experiment, where they had to use a helicopter simulator to navigate through courses with varying difficulty with the aim of flying through circles. While the subjects interacted with the simulator, we recorded their EEG in order to investigate whether their neural activity reflected their their performance in navigating the helicopter, as well as the varying difficulty of the simulator.

Figure 1 illustrates the concepts of the paradigm experiment.

1.1 Experimental setup

The experiment was constructed in the following way: A *trial* is defined as the time spent until reaching a circle, with a *block* consisting of nine trials. The subjects were instructed to take small breaks

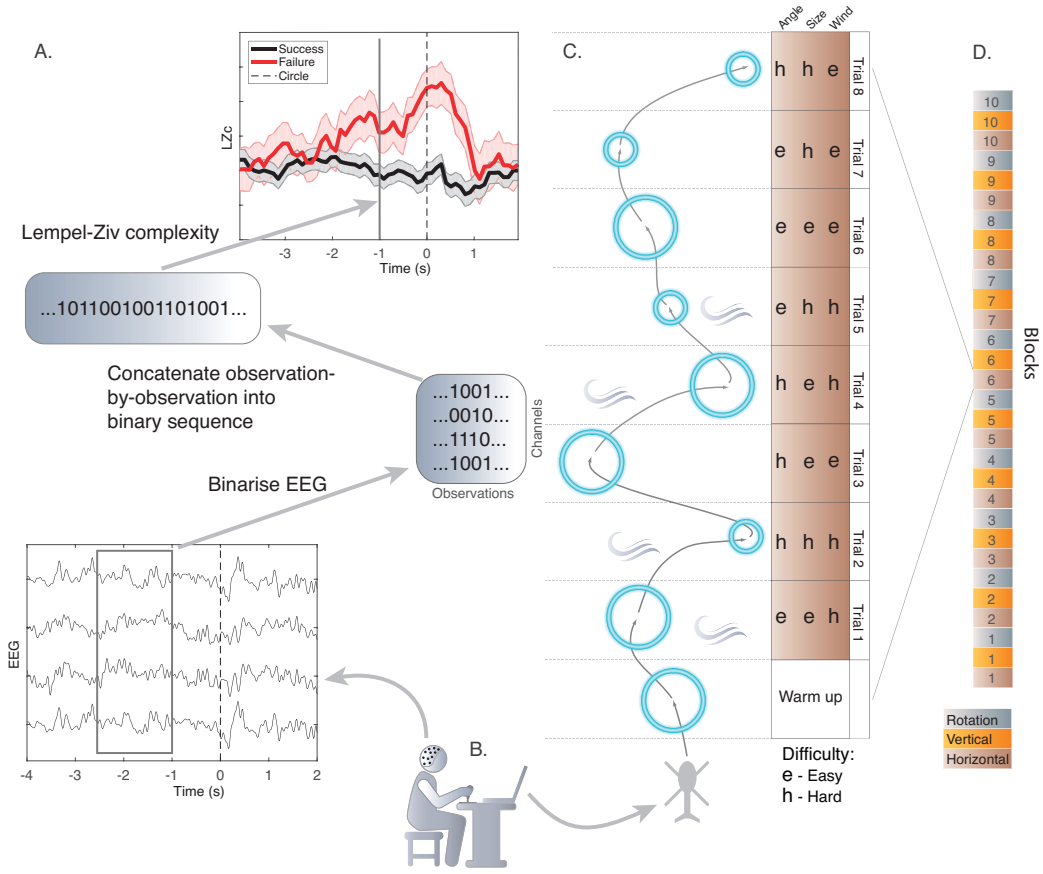


Figure 1: Experimental setup and EEG complexity pipeline. **A:** From EEG to LZc. EEG is recorded from a subject interacting with the helicopter simulator (**B**). In windows of 1,500 ms, the EEG is binarised based on the envelope amplitude of each channel. The binary matrix is "stretched" into a string observation-by-observation and given to the Lempel-Ziv algorithm. By using a moving window (with steps of 100 ms) and averaging across trials, the we can compare the difference in complexity leading up to a successful or failed navigation.
C: The object for the subject is to navigate a helicopter successfully through circles, where a trial is defined as the time leading to to reaching a circle. In total, the subject has to navigate a helicopter through 240 trials divided into blocks of eight trials. Each trial is affected by three independent types of difficulty, with an easy and a hard setting, and each block contains all eight difficulty combinations in random order. Trial 4, in the illustrated block, has a hard angle between the circles, the circle is large (easy) and there is turbulence from the hard wind difficulty.
D: In the helicopter simulator the subject has control the helicopter in three navigational modes, where the control scheme differs. The modes are played in a fixed order and repeated ten times.

between blocks, when needed. The subject navigated the helicopter in three different navigational *modes*, that alternated in a fixed order. The three navigational modes were then repeated ten times as can be seen in fig. 1, summing to a total of 270 trials per subject. The order of blocks and modes is illustrated on fig. 1**D**.

The helicopter simulator was designed specifically for this experiment, and featured three different and independent ways to wary the difficulty. Each difficulty had a hard and an easy setting, giving eight possible combinations of the three types of difficulty. Each block started with a "warm-up" trial and the remaining eight trials consisted of all posbiile difficulty combinations in random order. The three types of difficulty were: The *angle* difficulty determined the angle between the next circle

and the one just completed, meaning that a hard difficulty forces the subject to take sharper turns with the helicopter. The *size* difficulty determines the size of the circle the helicopter needs to be navigated through, where a hard difficulty requires the subject to have a better and smoother control over the helicopter. The *wind* difficulty decides whether or not there will be turbulence, which makes it harder to have a controlled navigation of the helicopter. This is also illustrated on fig. 1**C**.

1.2 Acquisition and pre-processing

During the experiment, we recorded electroencephalogram (EEG) from subjects using a 64 channel Biosemi system with active electrodes. All subsequent processing and analysis of the EEG was conducted digitally in Matlab.

Before analysing the EEG, we pre-processed it to remove unwanted noise in the recorded signals. First, the raw EEG was filtered using a 85 Hz lowpass and 50 Hz bandstop FIR filters to ensure linear phase and avoid phase distortion in the EEG. As we intended to measure the information sharing between brain areas, we opted not to employ digital re-referencing, to avoid inducing signals between distant electrodes. Therefore the only interactions between electrodes were with the CMS and DRL electrodes, which are placed occipitally in the Biosemi system.

We wanted to avoid signals of non-cortical origin (e.g. from eye or body movements) inducing the same information in multiple electrodes and thereby false positives in our analysis. We therefore employed independent component analysis (ICA), followed by dipole-fitting and automatic classification of the ICs using plug-ins for the EEGLAB toolbox, to help identify these non-cortical sources and remove them from the EEG (Delorme and Makeig, 2004; Oostenvelt et al., 2003; Frølich et al., 2015).

The experiment contained more movement than normally seen in experiments where EEG is recorded. EEG can be sensitive to movement artefacts, which in our experiment might create false positives for difficult trials, where the subject might unintentionally move their entire body. We therefore implemented a very aggressive preprocessing, only keeping ICs that the toolbox classified as most likely to be cortical. This resulted in a high number of rejected ICs.

Finally, we used surface Laplacian, with a neighbour count of eight to reduce the influence of volume conduction (BCILAB plug-in for EEGLAB; Kothe and Makeig (2013)).

1.3 Complexity analysis of EEG

In order to analyse EEG complexity, we employed Lempel-Ziv complexity implemented the same manner as in Andrillon et al. (2016). Briefly, LZc measures the complexity of a signal by estimating how much it can be compressed (Lempel and Ziv, 1976). The analysis is based on the algorithm used in known file lossless compression programs such as WinZip. LZc can be normalised such that it takes values between 0 (fully compressible, i.e., predictable signal) and 1 (minimally compressible). Previous studies have shown that LZc (or variants of it) can track levels of consciousness in patients as well as healthy subjects under anaesthesia and during sleep (Casali et al., 2013; Schartner et al., 2015; Andrillon et al., 2016; Schartner et al., 2017). The aim of this study was to see if this measure would also be able to track consciousness in awake subjects, while dealing with a task of varying difficulty.

It is worth noting that this analysis is unsupervised, meaning that it requires no models to be fitted to the data, and therefore there is no risk of over-fitting a model to observation noise.

For the analyses presented here, we use a subset of eight electrodes from the occipital-parietal area. This subset was selected based on tests from pilot experiments on an early version of the helicopter simulator.

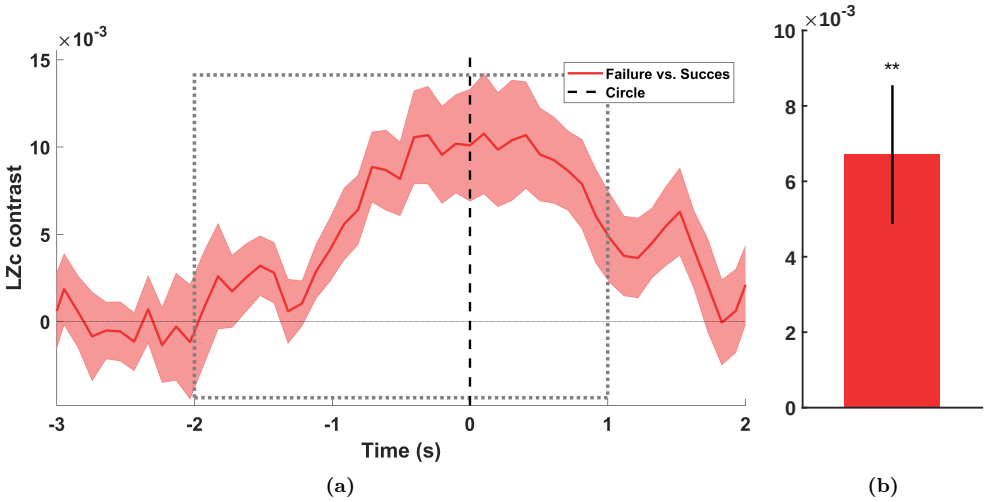


Figure 2: (a): LZc time-locked to the time point when the helicopter reached a circle (defined as t = 0 s) for the contrast between failed and successful trials. The LZc contrast were calculated by subtracting the average of successful trials from the average of failed trials for each subject. The contrast was then averaged across subjects, with shaded areas signify SEM. Note that two groups of trials was not balanced with respect to the number of trials. The dashed grey square indicates the temporal ROI used to average into single LZc values for each trial, which have been averaged across subjects in **(b)**. Error bars signify the standard error of the over the subjects mean.

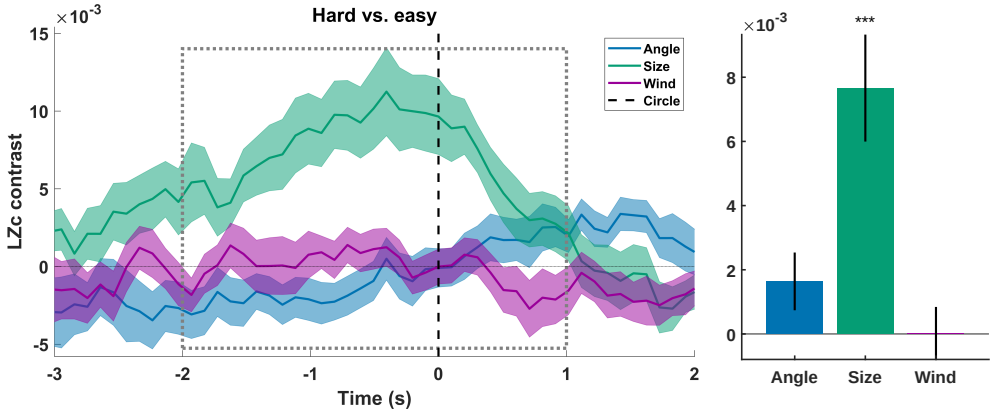
2 Results

With our experiment we wanted to investigate the capacity of EEG complexity to reflect neural changes leading up to failures while navigating a helicopter. To this end, we time-locked the LZc to when the helicopter reached a circle, and averaged across trials. On the subject level, we could see a clear differences in the LZc between successful and failed trials, but we also identified a strong subject-specific difference in the LZc (see figures S1 and S2). Due to our experimental paradigm, we were unable to remove subject-specific variations in the LZc, which we solved by looking at condition contrasts within each subject.

For each subject we calculated the average LZc for failed trials and subtracted them by the average LZc from successful trials, resulting in the graph shown on fig. 2a. From this contrast it can be seen that the EEG on average is more complex in trials where the subject is about to fail a navigation compared to successful trials. It can also be seen that there is a continued contrast after the circle has been reached, which is likely driven by the 1,500 ms width of the window used to calculate the LZc, which ends in the time point it is designated to. I.e. the LZc designated to t = 1,000 ms is calculated from the EEG between -500 ms to 1,000 ms, relative to reaching a circle.

To test the significance of this contrast we selected a temporal region of interest (ROI) from -2,000 ms to 1,000 ms. For each trial we averaged the LZc in this ROI, excluding LZc samples influenced by the ERP generated from the previous trial, as discussed in section 1.3, and calculated the subject-averages for the two trial subsets. Using a paired t-test showed a that the LZc in the ROI was significantly higher in failed trails, than in successful trials ($p = 0.001$).

In a real-world scenario, a flight-simulator would in most cases not need neural measures to register, when a navigational failure was made. As in the helicopter-simulator created for this experiment, failures could simply be recorded, when the subject failed to complete a task, such as flying through a circle. However, it is much harder for a simulator to register, when the user is struggling with navigating the helicopter. To investigate this, we looked at how the LZc varied with the changes in



(a) (b)

Figure 3: (a): LZc time-locked to the time point when the helicopter reached a circle (defined as t = 0 s) for three contrast between difficult and easier trials. The LZc contrasts were calculated by subtracting the average of easy trials from the average of hard trials for each subject. The contrasts were then averaged across subjects, with shaded areas signify SEM. The *Angle*, *Size*, and *Wind* contrasts are calculated using the hard and easy difficulty and the members of the two groups are therefore balanced. The dashed grey square indicates the temporal ROI used to average into single LZc values for each trial, which have been averaged across subjects in **(b)**. Error bars signify the standard error of the over the subjects mean.

the three types of difficulty.

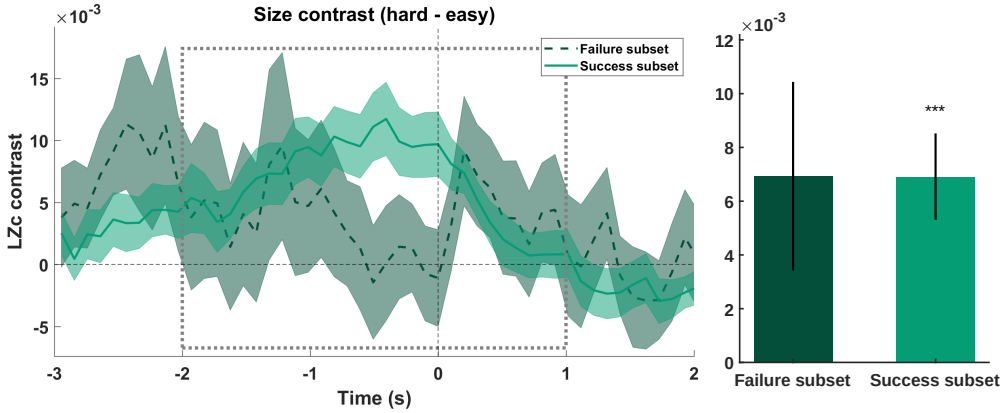
We conducted a repeated measures ANOVA with the LZc in the temporal ROI as the response variable and a within-subject design. As within-subject predictor variables we included the three difficulty types and the navigational mode, resulting in a 2-by-2-by-2-by-3 ANOVA. The ANOVA showed that the subjects, the *Size* difficulty, and the the navigational mode had a significant effects on the LZc variation ($p_{subject} < 0.001,\ p_{size} < 0.001,\ p_{mode} = 0.0191$), whereas the *Angle* and *Wind* difficulties had low or no effect on the variation in LZc ($p_{angle} = 0.077,\ p_{wind} = 0.977$).

In a manner similar to the failure contrast, we also investigated the influence of the types of difficulty by calculating contrasts between the trials with a hard and an easy setting. For example, for the *Angle* difficulty we subtracted the average LZc across all trials with an easy angle from the average of trials with a hard angle. Figure 3 supports the conclusion of the ANOVA with similar p-values ($p_{angle} = 0.077,\ p_{size} < 0.001,\ p_{wind} = 0.977$).

Comparing the failure contrast curve on fig. 2a with the *Size* contrast in fig. 3a, they can be seen to be were similar. This had us asking the question whether the contrast between hard and easy *Size* difficulty was simply because subjects failed the hard trials, so the LZc only captured neural changes due to upcoming failure, and not also struggles due to high difficulty. We therefore divided the trials into a subset of failed trials, and a subset of successful trials, and calculated the *Size* difficulty contrast for both subsets.

Figure 4 shows that, looking only at trials with successful navigations, there is still a significantly higher LZc when the *Size* difficulty contrast is hard. This is evidence that LZc not only reflects externally measurable events such as navigational failures, but also reflects more endogenous processes, such as when a user successfully completes a task, but struggles with it.

We were also interested in seeing whether the there was a change in LZc over time, which could reflect a learning process. We therefore include time, represented by the block number, as a predictive variable in the repeated measures ANOVA. This showed time to have a significant effect on LZc ($p < 0.001$), as well as having a significant interaction with the navigational mode ($p = 0.002$). We therefore investigated the difference of LZc in the temporal ROI between the first and the last block



(a) (b)

Figure 4: Contrasting difficulty in the two subsets of failed or successful trials. **(a):** Size difficulty contrast (hard vs. easy) is calculated in the same way as in fig. 3, but only for the *Failed* or the *Successful* subset of trials. Note that the trial count for *Failure* subset group is lower, and the average is therefore more noisy. Additionally, the low trial count might be compounded by the fact that short trials were susceptible to the removal of LZc samples influenced from the ERP of prior trial as discussed in section 1.3, which might explain the sudden change after $t = 0$. The dashed grey square indicates the temporal ROI used to average into single LZc values for each trial, which have been averaged across subjects in **(b)**. Error bars signify the standard error of the over the subjects mean.

for each navigational mode.

Figure 5 shows that all three navigational modes have a drop in complexity between the first and last block. The contrast is seen to be strongest for the Horizontal mode, and it is also the only mode to have a significant drop in complexity as measured be a paried t-test ($p = 0.003$). The Horizontal mode was also reported by subjects to be the easiest in post experiment questionnaires, as well as the mode were most quick to learn, as reported by self-scored learning curves.

3 Conclusion

Though not being able to distinguish on a single single trial level, we have shown that on average the EEG complexity, is higher in the seconds before missing circles compared to when they successfully navigated the helicopter through them.

All together, these results show that the EEG complexity shows promise in tracking the performance of persons mental workload while training a new task, such as learning to fly a helicopter simulator.

References

Andrillon, T., Poulsen, A. T., Hansen, L. K., Léger, D., and Kouider, S. (2016). Neural Markers of Responsiveness to the Environment in Human Sleep. *The Journal of Neuroscience*, 36(24):6583–6596.

Casali, a. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M.-a., Laureys, S., Tononi, G., and Massimini, M. (2013). A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. *Science Translational Medicine*, 5(198):1–10.

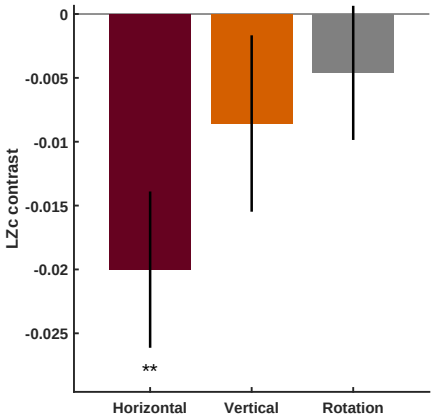


Figure 5: Change in LZc over time for three modes of navigation. For each trial the average LZc is calculated for the two seconds prior and 1 second subsequent to the helicopter reaching a circle. Contrasts are calculated by subtracting trials of the last block from trials from the first block, where the paired trials have the same configuration of difficulties (*Angle*, *Size* and *Wind*). Only includes contrasts from block pairs, where both trials were successful. Averaged across subjects with bars denoting SEM. Statistical significance was calculated using uncorrected paired t-tests.

Delorme, A. and Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

Frølich, L., Andersen, T. S., and Mørup, M. (2015). Classification of independent components of EEG into multiple artifact classes. *Psychophysiology*, 52(1):32–45.

Kothe, C. A. and Makeig, S. (2013). BCILAB: a platform for brain-computer interface development. *Journal of neural engineering*, 10(5):056014.

Lempel, A. and Ziv, J. (1976). On the Complexity of Finite Sequences. *IEEE Transactions on Information Theory*, 22(1):75–81.

Oostenvelt, R., Delorme, A., and Makeig, S. (2003). DIPFIT: Equivalent dipole source localization of independent components.

Schartner, M., Seth, A., Noirhomme, Q., Boly, M., Bruno, M. A., Laureys, S., and Barrett, A. (2015). Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia. *PLoS ONE*, 10(8):1–21.

Schartner, M. M., Pigorini, A., Gibbs, S. A., Arnulfo, G., Sarasso, S., Barnett, L., Nobili, L., Massimini, M., Seth, A. K., and Barrett, A. B. (2017). Global and local complexity of intracranial EEG decreases during NREM sleep. *Neuroscience of Consciousness*, (September 2016):1–12.

Supplementary information for

Tracking difficulty in a helicopter simulator: EEG complexity as a marker for mental workload

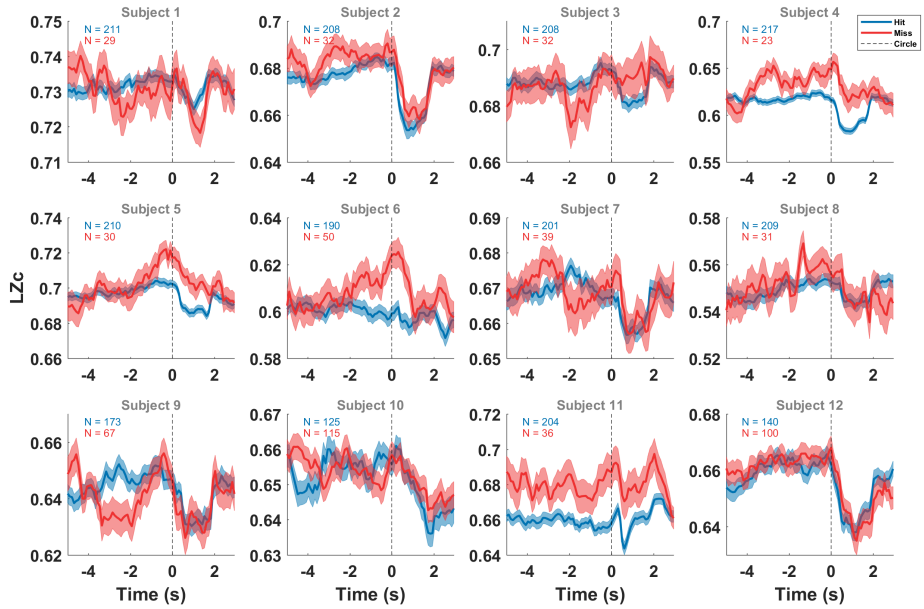


Figure S1: Average LZc for subject 1-12, synchronised to the time point when helicopter reached a circle. The LZc is averaged across trials where helicopter went through circle ("Hit"), and when helicopter went past circle ("Missed"). Shaded bars signify SEM over trials.

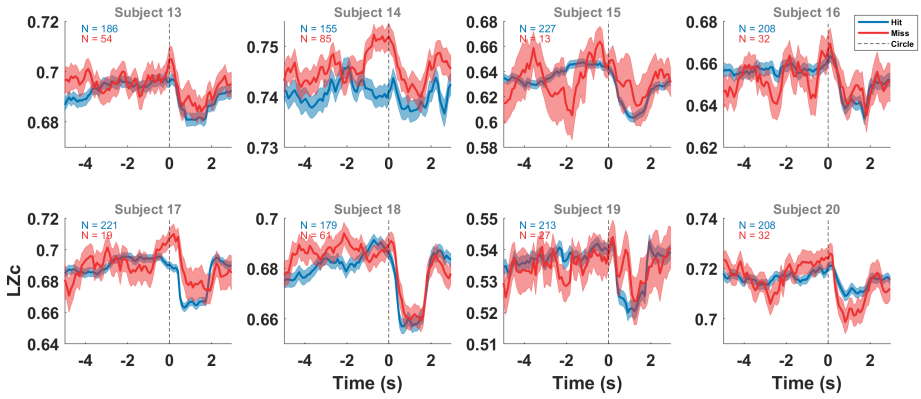


Figure S2: Average LZc for subject 13-20, synchronised to the time point when helicopter reached a circle. The LZc is averaged across trials where helicopter went through circle ("Hit"), and when helicopter went past circle ("Missed"). Shaded bars signify SEM over trials.

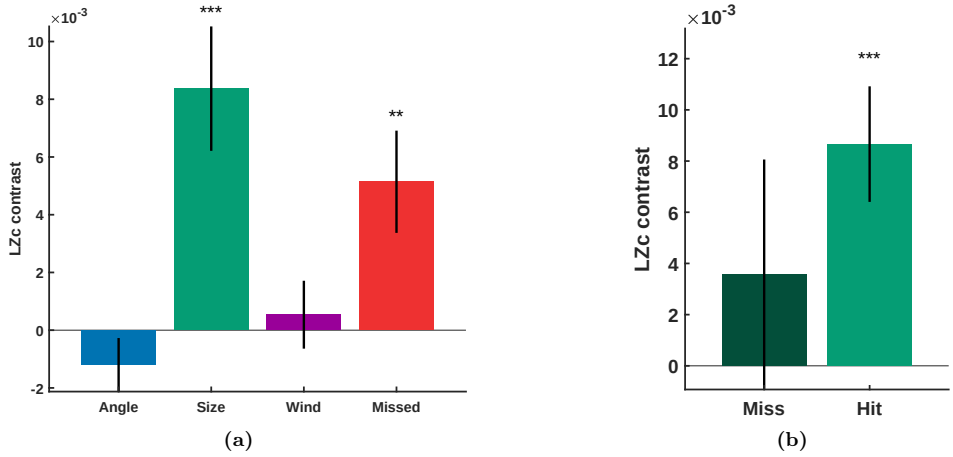


Figure S3: Predictive capacity of LZc. Contrasts over the hard vs. easy difficulty using only EEG recorded prior to reaching a circle. Note that the trial count for *Missed* group is lower, and the average is therefore more noisy. Note that the LZc is calculated in 1.5 s windows ending in the time point designated for the LZc value. The LZc presented here is therefore not based on EEG recorded subsequent to reaching a circle.

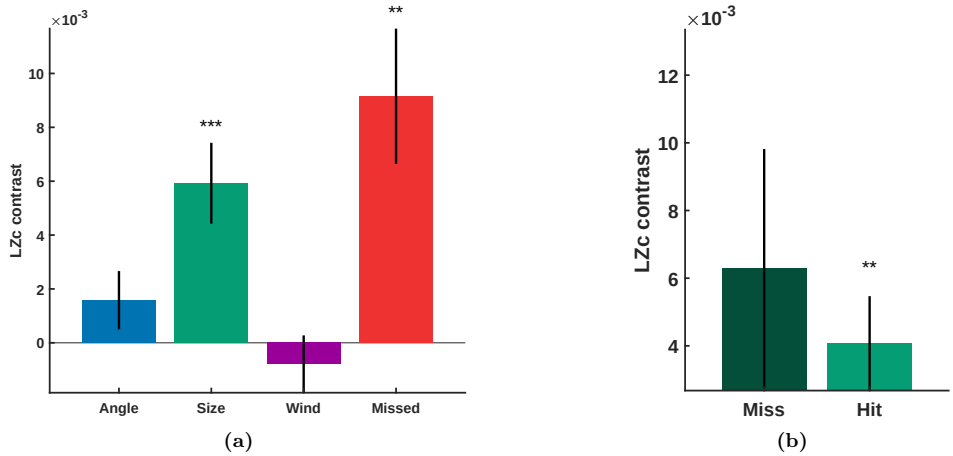


Figure S4: LZc only for the 1 second after reaching the circle. Contrasts over the hard vs. easy difficulty using only EEG recorded subsequent to reaching a circle. Note that the trial count for *Missed* group is lower, and the average is therefore more noisy. Note that the LZc is calculated in 1.5 s windows and therefore include EEG prior to reaching the circle.

APPENDIX E

Tracking difficulty in a helicopter simulator: EEG complexity as a marker for mental workload

Authors: **Andreas Trier Poulsen**, Jean-maurice Leonetti, Lars Kai Hansen, and Sid Kouider.

Status: Abstract submitted for conference.

Tracking difficulty in a helicopter simulator: EEG complexity as a marker for mental workload

Electroencephalography (EEG) is an intricate multi-dimensional measure, and there are many approaches to analyse EEG or compare it with the stimuli subjects are interacting with.

Recent research has shown that the complexity of EEG signals is correlated with the levels of consciousness in comatose patients (Casali et al., 2013) as well as healthy subjects under anaesthesia and during sleep (Schartner et al.,2015; Andrillon et al., 2016; Schartner et al.,2017).

In this study we investigate whether EEG complexity (LZc), as captured by the Lempel-Ziv algorithm (Lempel and Ziv, 1976), can be used in fully aware, healthy people as an index of how focused they are on a given task.

Twenty subjects (hereof 10 were female) were recruited for an experiment, where they had to use a helicopter simulator to navigate through courses with varying difficulty with the aim of flying through circles. While the subjects interacted with the simulator, we recorded their EEG in order to investigate whether their neural activity reflected their performance of navigating the helicopter, as well as the varying difficulty of the simulator.

This paradigm contained a higher degree of movement than normally seen in experiments where EEG is recorded. EEG can be sensitive to movement artefacts, which in our experiment might create false positives for difficult trials, due to the subject unintentionally moving their entire body. We therefore implemented an aggressive preprocessing using independent component analysis (ICA), followed by dipole-fitting and automatic classification of the ICs using plug-ins for the EEGLAB toolbox, to help identify non-cortical sources and remove them from the EEG (Delorme and Makeig, 2004; Oostenvelt et al., 2003; Frølich et al., 2015).

The helicopter simulator was designed for this experiment and featured three modes with different ways to navigate the helicopter, as well as three different ways to wary the difficulty. By contrasting each of these difficulty types, we could investigate whether LZc was able to identify trials with high difficulty, where the subjects were assumed to struggle more.

Furthermore, by contrasting the trials, where subjects were successful in navigating the helicopter, with failed trials, we could identify moments of high mental workload and investigate how well LZc was able to track these moments.

Though not distinguishable on a single-trial level, subjects showed significantly higher complexity on average in the seconds before failing a trial compared to when they successfully navigated the helicopter through the circles (see figure 1). Additionally, a significant drop in complexity was measured in the navigational mode, which subjects reported as being the easiest. This mode was presumably the one they improved the most in, thereby reaching a plateau in their improvement early on. This could result in a decrease in focus, reflected in the decreased LZc.

The difficulty type that obtained the highest LZc contrast between easy and hard trials, was also the difficulty type that was the most influential in whether subjects failed a trial. This was also the case when calculating this difficulty contrast only on successful trials, which indicates that the LZc not only captures neural changes up to a failure, but also when a subject is struggling during a hard but successful trial.

That LZc is able to distinguish moments of varying workload consistently across subjects, suggests that EEG complexity is a viable candidate as an index of mental workload.

Andrillon, T., Poulsen, A. T., Hansen, L. K., Léger, D., and Kouider, S. (2016). Neural Markers of Responsiveness to the Environment in Human Sleep. The Journal of Neuroscience, 36(24):6583–6596.

Casali, a. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M.-a., Laureys, S., Tononi, G., and Massimini, M. (2013). A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. Science Translational Medicine, 5(198):1–10.

Delorme, A. and Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134(1):9–21.

Frølich, L., Andersen, T. S., and Mørup, M. (2015). Classification of independent components of EEG into multiple artifact classes. Psychophysiology, 52(1):32–45.

Lempel, A. and Ziv, J. (1976). On the Complexity of Finite Sequences. IEEE Transactions on Information Theory, 22(1):75–81.

Oostenvelt, R., Delorme, A., and Makeig, S. (2003). DIPFIT: Equivalent dipole source localization of independent components.

Schartner, M., Seth, A., Noirhomme, Q., Boly, M., Bruno, M. A., Laureys, S., and Barrett, A. (2015). Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia. PLoS ONE, 10(8):1–21.

Schartner, M. M., Pigorini, A., Gibbs, S. A., Arnulfo, G., Sarasso, S., Barnett, L., Nobili, L., Massimini, M., Seth, A. K., and Barrett, A. B. (2017). Global and local complexity of intracranial EEG decreases during NREM sleep. Neuroscience of Consciousness, (September 2016):1–12

Bibliography

Abasolo, D., Hornero, R., Gomez, C., Garcia, M., and Lopez, M. (2006). Analysis of EEG background activity in Alzheimer's disease patients with Lempel-Ziv complexity and central tendency measure. *Medical Engineering & Physics*, 28(4):315–322.

Aboy, M., Hornero, R., Abásolo, D., and Álvarez, D. (2006). Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Transactions on Biomedical Engineering*, 53(11):2282–2288.

Andrillon, T., Poulsen, A. T., Hansen, L. K., Léger, D., and Kouider, S. (2016). Neural Markers of Responsiveness to the Environment in Human Sleep. *The Journal of Neuroscience*, 36(24):6583–6596.

Ashburner, J., Barnes, G., Chen, C., Daunizeau, J., Flandin, G., Friston, K., Kiebel, S., Kilner, J., Litvak, V., and Moran, R. (2014). SPM12 manual. *London: Wellcome Trust*.

Bordwell, D. (2002). Intensified continuity: visual style in contemporary American film. *Film Quarterly*, 55(3):16–28.

Britz, J., Van De Ville, D., and Michel, C. M. (2010). BOLD correlates of EEG topography reveal rapid resting-state network dynamics. *NeuroImage*, 52(4):1162–1170.

Buzsáki, G., Anastassiou, C. a., and Koch, C. (2012). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6):407–420.

Casali, a. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M.-a., Laureys, S., Tononi, G., and Massimini, M. (2013). A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. *Science Translational Medicine*, 5(198):1–10.

Cichy, R. M., Pantazis, D., and Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, 17(3):455–462.

Cohen, S. S., Madsen, J., Touchan, G., Robles, D., Lima, S. F. A., Henin, S., and Parra, L. C. (2018). Neural engagement with online educational videos predicts learning performance for individual students. *bioRxiv*.

Cohen, S. S. and Parra, L. C. (2016). Memorable Audiovisual Narratives Synchronize Sensory and Supramodal Neural Responses. *eNeuro*, 3(6):1–11.

Cortese, A., Amano, K., Koizumi, A., Kawato, M., and Lau, H. (2016). Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nature Communications*, 7:13669.

Dehaene, S. and Changeux, J. P. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, 70(2):200–227.

Delorme, A. and Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J. J., Ding, M., and Poeppel, D. (2017). Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Current Biology*, 27(9):1375–1380.

Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., and Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nature Communications*, 5:1–9.

Dmochowski, J. P., Sajda, P., Dias, J., and Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? *Frontiers in human neuroscience*, 6(May):112.

Frølich, L., Andersen, T. S., and Mørup, M. (2015). Classification of independent components of EEG into multiple artifact classes. *Psychophysiology*, 52(1):32–45.

Hardoon, D. R., Szedmak, S., and Shawe-taylor, J. (2004). Canonical correlation analysis; An overview with application to learning methods. *Neural computation*, 16(12):2639–2664.

Hari, R. and Puce, A. (2017). *MEG-EEG primer*, volume 1. Oxford University Press.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., and Heeger, D. J. (2008). Neurocinematics: The Neuroscience of Film. *Projections*, 2(1):1–26.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. *Science*, 303(5664):1634–1640.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., and Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87:96–110.

Hotelling, H. (1936). Relations Between Two Sets of Variates. *Biometrika*, 28(3/4):321.

Iber, C., Ancoli-Israel, S., Chesson, A. J., and Quan, S. (2007). *The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications*. American Academy of Sleep Medicine, Westchester.

Jensen, B. S., Larsen, J., Jensen, K., Larsen, J. E., and Hansen, L. K. (2010). Predictability of Mobile Phone Associations. *Inter. Workshop on Mining Ubiquitous and Social Environments*.

Jing Hu, Jianbo Gao, and Principe, J. (2006). Analysis of Biomedical Signals by the Lempel-Ziv Complexity: the Effect of Finite Data Size. *IEEE Transactions on Biomedical Engineering*, 53(12):2606–2609.

Johannes, S., Münte, T. F., Heinze, H. J., and Mangun, G. R. (1995). Luminance and spatial attention effects on early visual processing. *Cognitive Brain Research*, 2:189–205.

Kamronn, S. and Poulsen, A. (2013). *Machine Learning for Social EEG-A Bayesian approach to correlated component analysis and recording simultaneous multiple subject EEG*. Master thesis, Technical University of Denmark.

Kamronn, S., Poulsen, A. T., and Hansen, L. K. (2015). Multiview Bayesian Correlated Component Analysis. *Neural computation*, 27(10):1–24.

Keil, J., Muller, N., Ihssen, N., and Weisz, N. (2012). On the Variability of the McGurk Effect: Audiovisual Integration Depends on Prestimulus Brain States. *Cerebral Cortex*, 22(1):221–231.

Khaligh-Razavi, S. M. and Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLoS Computational Biology*, 10(11).

Khanna, A., Pascual-Leone, A., and Farzan, F. (2014). Reliability of resting-state microstate features in electroencephalography. *PLoS ONE*, 9(12):1–21.

Khanna, A., Pascual-Leone, A., Michel, C. M., and Farzan, F. (2015). Microstates in resting-state EEG: Current status and future directions. *Neuroscience & Biobehavioral Reviews*, 49:105–113.

Ki, J. J., Kelly, S. P., and Parra, L. C. (2016). Attention Strongly Modulates Reliability of Neural Responses to Naturalistic Narrative Stimuli. *Journal of Neuroscience*, 36(10):3092–3101.

Kidmose, P., Looney, D., Jochumsen, L., and Mandic, D. P. (2013). Ear-EEG from generic earpieces: A feasibility study. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 543–546.

King, J. R., Pescetelli, N., and Dehaene, S. (2016). Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. *Neuron*, 92(5):1122–1134.

King, J.-R., Sitt, J. D., Faugeras, F., Rohaut, B., El Karoui, I., Cohen, L., Naccache, L., and Dehaene, S. (2013). Information sharing in the brain indexes consciousness in noncommunicative patients. *Current biology : CB*, 23(19):1914–9.

Klami, A. (2013). Bayesian Canonical Correlation Analysis. *Journal of Machine Learning Research*, 14:965–1003.

Ko, L.-W., Komarov, O., Hairston, W. D., Jung, T.-P., and Lin, C.-T. (2017). Sustained Attention in Real Classroom Settings: An EEG Study. *Frontiers in Human Neuroscience*, 11(July):1–10.

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). The neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience*, in press(5):307–321.

Kothe, C. A. and Makeig, S. (2013). BCILAB: a platform for brain-computer interface development. *Journal of neural engineering*, 10(5):056014.

Kouider, S., Andrillon, T., Barbosa, L. S., Goupil, L., and Bekinschtein, T. A. (2014). Inducing task-relevant responses to speech in the sleeping brain. *Current Biology*, 24(18):2208–2214.

Kouider, S., Stahlhut, C., Gelskov, S. V., Barbosa, L. S., Dutat, M., de Gardelle, V., Christophe, A., Dehaene, S., and Dehaene-Lambertz, G. (2013). A neural marker of perceptual consciousness in infants. *Science (New York, N.Y.)*, 340(6130):376–80.

Lahnakoski, J. M., Glerean, E., Jääskeläinen, I. P., Hyönä, J., Hari, R., Sams, M., and Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*, 100:316–24.

Langer, N., Ho, E. J., Alexander, L. M., Xu, H. Y., Jozanovic, R. K., Henin, S., Petroni, A., Cohen, S., Marcelle, E. T., Parra, L. C., Milham, M. P., and Kelly, S. P. (2017). A resource for assessing information processing in the developing brain using EEG and eye tracking. *Scientific data*, 4:170040.

Lehmann, D., Ozaki, H., and Pal, I. (1987). EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. *Electroencephalography and Clinical Neurophysiology*, 67(3):271–288.

Lehmann, D., Strik, W. K., Henggeler, B., Koenig, T., and Koukkou, M. (1998). Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. Visual imagery and abstract thoughts. *International Journal of Psychophysiology*, 29(1):1–11.

Lempel, A. and Ziv, J. (1976). On the Complexity of Finite Sequences. *IEEE Transactions on Information Theory*, 22(1):75–81.

Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., and Wass, S. (2017). Speaker gaze increases information coupling between infant and adult brains. *Proceedings of the National Academy of Sciences*, page 201702493.

Li, Y., Tong, S., Liu, D., Gai, Y., Wang, X., Wang, J., Qiu, Y., and Zhu, Y. (2008). Abnormal EEG complexity in patients with schizophrenia and depression. *Clinical Neurophysiology*, 119(6):1232–1241.

Lopes da Silva, F. (2013). EEG and MEG: Relevance to neuroscience. *Neuron*, 80(5):1112–1128.

Mesgarani, N. and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397):233–236.

Michel, C. M. (2009). *Electrical neuroimaging*. Cambridge University Press.

Mirkovic, B., Debener, S., Jaeger, M., and De Vos, M. (2015). Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. *Journal of Neural Engineering*, 12(4):046007.

Müller, M. M., Andersen, S., Trujillo, N. J., Valdés-Sosa, P., Malinowski, P., and Hillyard, S. a. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 103(38):14250–4.

Murray, M. M., Brunet, D., and Michel, C. M. (2008). Topographic ERP analyses: A step-by-step tutorial review. *Brain Topography*, 20(4):249–264.

Naci, L., Sinai, L., and Owen, A. M. (2015). Detecting and interpreting conscious experiences in behaviorally non-responsive patients. *NeuroImage*.

Norton, J. J. S., Sup, D., Woo, J., Lee, W., Kwon, O., Won, P., and Jung, S.-y. (2015). Soft , curved electrode systems capable of integration on the auricle as a persistent brain – computer interface. *PNAS*.

Oostenvelt, R., Delorme, A., and Makeig, S. (2003). DIPFIT: Equivalent dipole source localization of independent components.

Parra, L. C. (2018). Multi-set Canonical Correlation Analysis simply explained. *arXiv*.

Parra, L. C., Haufe, S., and Dmochowski, J. P. (2018). Correlated Components Analysis — Extracting Reliable Dimensions in Multivariate Data. *arXiv*, pages 1–40.

Pascual-Marqui, R., Michel, C., and Lehmann, D. (1995). Segmentation of brain electrical activity into microstates: model estimation and validation. *Biomedical Engineering, IEEE Transactions on*, 42(7):658–665.

Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., and Hansen, L. K. (2017). EEG in the classroom: Synchronised neural recordings during video presentation. *Scientific Reports*, 7:1–9.

Poulsen, A. T., Pedroni, A., Langer, N., and Hansen, L. K. (2018). Microstate EEGlab toolbox: An introductory guide. *bioRxiv*, pages 1–30.

Rokach, L. and Maimon, O. (2005). Clustering Methods. In *Data Mining and Knowledge Discovery Handbook*, pages 321–352. Springer US, New York.

Schartner, M., Seth, A., Noirhomme, Q., Boly, M., Bruno, M. A., Laureys, S., and Barrett, A. (2015). Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia. *PLoS ONE*, 10(8):1–21.

Schartner, M. M., Pigorini, A., Gibbs, S. A., Arnulfo, G., Sarasso, S., Barnett, L., Nobili, L., Massimini, M., Seth, A. K., and Barrett, A. B. (2017). Global and local complexity of intracranial EEG decreases during NREM sleep. *Neuroscience of Consciousness*, (September 2016):1–12.

Stopczynski, A., Stahlhut, C., Larsen, J. E., Petersen, M. K., and Hansen, L. K. (2014a). The Smartphone Brain Scanner: A Portable Real-Time Neuroimaging System. *PLoS ONE*, 9(2):e86733.

Stopczynski, A., Stahlhut, C., Petersen, M. K., Larsen, J. E., Jensen, C. F., Ivanova, M. G., Andersen, T. S., and Hansen, L. K. (2014b). Smartphones as pocketable labs: visions for mobile brain imaging and neurofeedback. *International journal of psychophysiology*, 91(1):54–66.

Tibshirani, R. and Walther, G. (2005). Cluster Validation by Prediction Strength. *Journal of Computational and Graphical Statistics*, 14(3):511–528.

Van De Ville, D., Britz, J., and Michel, C. M. (2010). EEG microstate sequences in healthy humans at rest reveal scale-free dynamics. *Proceedings of the National Academy of Sciences*, 107(42):18179–18184.

Viola, F. C., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., and Debener, S. (2009). Semi-automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology*, 120(5):868–877.

Wakeman, D. G. and Henson, R. N. (2015). A multi-subject, multi-modal human neuroimaging dataset. *Scientific data*, 2:150001.

Wong, D. D. E., Fuglsang, S. A., Hjortkjær, J., Ceolini, E., Slaney, M., and de Cheveigné, A. (2018). A Comparison of Temporal Response Function Estimation Methods for Auditory Attention Decoding. *bioRxiv*.

Yuan, H., Zotev, V., Phillips, R., Drevets, W. C., and Bodurka, J. (2012). Spatiotemporal dynamics of the brain at rest - Exploring EEG microstates as electrophysiological signatures of BOLD resting state networks. *NeuroImage*, 60(4):2062–2072.

Zhang, X. S., Roy, R. J., and Jensen, E. W. (2001). EEG complexity as a measure of depth of anesthesia for patients. *IEEE Transactions on Biomedical Engineering*, 48(12):1424–1433.